Probability and Statistics

3rd year degree honours course

Dr. Arnab Chakraborty

Associate Professor
Indian Statistical Institute, Kolkata

Formerly

Guest Faculty
Ramakrishna Mission Vidyamandira, Belur

◇

Guest Faculty
Ramakrishna Vivekananda University, Belur

◇

Assistant Professor
St. Xavier's College, Kolkata

Sarat Book Distributors
18-B, Shyamacharan De Street
Kolkata 700 073

সূচী

ছাত্রছাত্রীদের প্রতি

আমাদের বাংলায়-বোঝানো-ইংরাজি বই সিরিজের এই বইটা দ্বিতীয় সংস্করণে পা রাখল। আগের সংস্করণের ছাপার ভুলটুলগুলো যথাসম্ভব ঠিক করা হয়েছে এই সংস্করণে। যারা ভুলগুলো ধরতে সাহায্য করেছ, তাদের সকলকে অনেক অনেক ধন্যবাদ।
আমাদের এই সিরিজটা লেখা হচ্ছে সেইসব ছাত্রছাত্রীদের জন্য, যারা ইংরাজি মাধ্যমে পড়ছ, কিন্তু যাদের ইংরাজির চেয়ে বাংলাতে বুঝতেই বেশী সুবিধা হয়। এ বইতে যাবতীয় definition, theorem, ইত্যাদি ইংরাজিতে লেখা। দশ বছরের যাবতীয় প্রশ্নের বিশদ উত্তর আলোচনা করা হয়েছে। সমাধানের যে অংশটুকু তোমাকে লিখতে হবে, সেটা handwriting-এর মত করে দেওয়া আছে, ফাঁকে ফাঁকে বাংলায় বোঝানো। ঠিক যেন ক্লাসে মাষ্টারমশাই পড়াচ্ছেন, হাতের লেখা অংশগুলো বোর্ডে লেখা হচ্ছে, আর বাংলাগুলো মুখে বলছেন।
পুরো বইটা Day 1, Day 2,... এইভাবে দিনে দিনে ভাঙা রয়েছে। একেকটা দিনে আমরা একেকটা জিনিস শিখব। একটা day বলে যতটা দিয়েছি, সেটা কোচিং ক্লাসের দু ঘন্টার পড়া। একদিনে ওর চেয়ে বেশী পড়ার চেষ্টা করলে মাথা ঝিমঝিম করবে।
যেদিন যেটুকু পড়লে, সেদিনই সেই অংশের দশ বছরের প্রশ্নগুলো শিখে নেওয়া ভালো। বইটা সেইভাবেই লেখা। দশ বছরের যাবতীয় প্রশ্নের উত্তর আলোচনা করা আছে, এমন কি ভুল প্রশ্নগুলোকেও বাদ দেওয়া হয় নি। সেগুলো কেন ভুল, বা কী করে মেরামত করে নেওয়া যাবে, তাও আলোচনা করেছি।
Probability আর statistics-এ মজা পেতে গেলে কম্পিউটারে কাজ করাটা খুবই দরকার, কারণ এই বিষয়টা তৈরীই হয়েছে প্রচুর পরিমাণ data থেকে তথ্য আহরণের জন্য। অল্প কিছু সংখ্যা নিয়ে কাজ করলে তাই এর মজা পাওয়া কঠিন। যদিও আজ মানুষের ঘরে ঘরে কম্পিউটার (এমন কি হাতে হাতেও কম্পিউটার, কারণ স্মার্টফোন মানেই কম্পিউটার), তাও প্রাচীনপন্থী সিলেবাসে কম্পিউটারের স্থান সীমিত। এই দুয়ের ভারসাম্য রাখার জন্য কিছু program, video ইত্যাদি রেখেছি এই ওয়েবসাইটে--

`http://www.isical.ac.in/~arnabc/probstat/`

এগুলো চালানোর জন্য কোনো কম্পিউটারের বিদ্যা লাগবে না, টুকটাক ক্লিক করলেই হবে। স্মার্টফোন থেকেও করা যাবে। বইয়ের ভিতর বিভিন্ন জায়গায় এই ওয়েবসাইটের উল্লেখ দেখবে।
দশ বছরের প্রশ্নগুলো নিয়েছি বাজারে প্রচলিত সংকলনগুলো থেকে। তাই প্রশ্নের ভুলের দায় কতটা প্রশ্নকর্তার আর কতটা সংকলনকারীর, তা বলতে পারব না। এই বইটাকে যথাসম্ভব ত্রুটিমুক্ত রাখতে চেষ্টা করেছি। তাও কিছু ত্রুটি থেকে যাওয়া অসম্ভব নয়। কোনো ভুল চোখে পড়লে, বা অন্য কোনো কারণে যোগাযোগ করতে চাইলে email (arnabc74@gmail.com) বা ফোন (9231542600) ব্যবহার করতে পারো। এই নম্বরে Whatsapp-ও করতে পারো।

অর্ণব চক্রবর্তী

Chapter I
একেবারে গোড়ার কথা

DAY 1
কী ও কেন?

1.1 ঠিক যেন ম্যাজিক

হ্যাঁ, এ তো বোঝাই যাচ্ছে--চন্দ্রবিন্দুর চ, বেড়ালের তালব্য শ, রুমালের মা-- হল চশমা। কেমন, হল তো?
--হ য ব র ল (সুকুমার রায়)

এই বইয়ের বিষয়বস্তু হল statistics আর probability. প্রথম দর্শনে এই জিনিসদুটোকে দেখতে এতই অদ্ভুত যে বহু ছাত্রছাত্রী গোড়াতেই খাবি খেয়ে থেমে যায়। কী করছি, কেন করছি--সবই কেমন যেন গুবলেট পাকিয়ে যায়। তাই প্রথমেই statistics আর probability-র মূলমর্মটা কয়েকটা উদাহরণ দিয়ে বুঝে রাখি।

তোমরা সবাই লুডো খেলেছ নিশ্চয়ই? বোর্ডের উপর গুটি থাকে, আর ছক্কা চেলে চেলে সেটাকে এগিয়ে নিয়ে যেতে হয়। পথে সাপ, মই ইত্যাদি রোমহর্ষক জিনিসপত্র থাকে। আমরা এরকম একটা লুডো খেলা দিয়ে শুরু করব। এটা অবশ্য বাচ্চাদের লুডো নয়, দস্তুরমত অংকের লুডো। ফলে সাপ-মইয়ের মজাটা এখানে পাবে না, কিন্তু যেটা পাবে সেটা দেখে তোমার চোখ এক্ষুণি কপালে উঠবে। এখানে বোর্ডটা হবে গোটা $\mathbb{R}^2$, মানে ভাবতে পারো যেন গ্রাফের পাতার উপর খেলা হচ্ছে। গুটি একটাই, সেটা একটা point. গোড়াতে সেটা আছে $(0,0)$-তে। এবার একটা ছক্কা চাই। সাধারণ ছক্কার ছয়টা দিক থাকে। আমাদের ছক্কার মোটে চারটে দিক হলেই চলবে। ভাবতে পারো যেন চার টুকরো কাগজে $1,2,3,4$ লিখে ভালো করে মিশিয়ে যেকোনো একটা কাগজ চোখ বুঁজে টেনে নিচ্ছি। তাহলে বোর্ড হল, গুটি হল, ছক্কাও হল। এবার গুটি এগোনোর জন্য কিছু নিয়মকানুন লাগবে। ধরো এখন আছি (x,y)-তে। তাহলে ছক্কায় কত পড়লে কোথায় যাব, তার হদিশ রয়েছে এই টেবিলে--

ছক্কায় কত পড়লে...	... কোথায় যাব
1	$(\ 0.8x+0.1,\quad 0.8y+0.04\ )$
2	$(\ 0.5x+0.25,\quad 0.5y+0.4\ )$
3	$(\ 0.355(x-y)+0.266,\quad 0.355(x+y)+0.078\ )$
4	$(\ 0.355(x+y)+0.378,\quad 0.355(y-x)+0.434\ )$

এত সংখ্যা দেখে ঘাবড়ে যেও না (ঠিক যেমন সাপ লুডোর সাপগুলোকে দেখে আঁৎকে ওঠার কোনো কারণ নেই)। কয়েক দান খেলে নিয়ে খেলার নিয়মগুলো সড়গড় হয়ে নেওয়া যাক। ধরো আমরা লুডোটা খেলা শুরু করলাম। প্রথমে গুটিটা আছে $(0,0)$-তে, সেখানে একটা ফুট্‌কি আঁকলাম (Fig 1)। এবার ছক্কা চাললাম, মনে কর 2 পড়ল। টেবিলে বলেছে যে, 2 পড়লে আমাদের যেতে হবে $(\ 0.5x+0.25,\quad 0.5y+0.4\ )$, যেখানে (x,y) হল বর্তমান অবস্থান। তার মানে আমাদের বেলায় $x=0, y=0$. তাই আমাদের যেতে হবে এইখানে--

$$(0.5\times 0+0.25, 0.5\times 0+0.4)=(0.25,0.4).$$

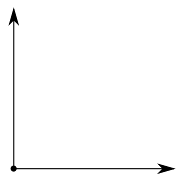


Fig 1

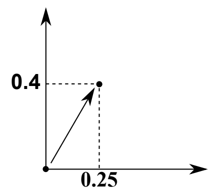


Fig 2

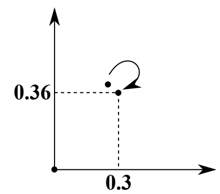


Fig 3

সেখানে আরেকটা ফুট্‌কি বসালাম (Fig 2)। আবার ছক্কা চালি, এবার ধরো পেলাম 1, ফলে গুটির পরবর্তী অবস্থান পাওয়া যাবে টেবিলের 1 নম্বর লাইনের ফর্মুলা থেকে, মানে ($0.8x + 0.1, \quad 0.8y + 0.04$). এখন রয়েছি $(x, y) = (0.25, 0.4)$-এ, সুতরাং এবার যাব এইখানে--

$$(0.8 \times 0.25 + 0.1, 0.8 \times 0.4 + 0.04) = (0.3, 0.36).$$

সেখানে ফের ফুট্‌কি বসালাম (Fig 3)। এইভাবে যদি বার বার খেলতেই থাকি তবে ফুট্‌কির সংখ্যা ক্রমশঃই বাড়বে। এবার বলো তো, যদি বহুবার খেলি তবে সবগুলো ফুট্‌কি মিলে কিরকম আকার নেবে? সম্পূর্ণ এলোমেলো কিছু একটা? নাকি একটা বৃত্তের মত? নাকি অন্য কিছু? উত্তরটা অতি অদ্ভুত, ধাপে ধাপে দেখাই। 100 বার খেলার পর চেহারাটা হয়েছে Fig 4-এর মত, যার মধ্যে তেমন কোনো প্যাটার্ণ চোখে পড়ছে না। 1000 বারের পর Fig 5-এর মত, এবার কিন্তু আবছা ছবির মত কিছু একটা ফুটে উঠেছে। যদি 10000 বার খেলি তবে (ও মা!) চমৎকার একটা পাতার ছবি হয়ে গেল কী করে (Fig 6)? এই জিনিসটা এতই বিস্ময়কর যে নিজের হাতে করে দেখলে মজা পাবে। অবশ্য 10000 বার কাগজ টেনে অংক কষে ফুট্‌কি আঁকতে গেলে মজা ওখানেই বেরিয়ে যাবে। তাই সেই কাজটা কম্পিউটার দিয়ে করার মত একটা প্রোগ্রাম লিখে রেখেছি যেটা চাইলে তুমিও চালাতে পারো বাড়ির কম্পিউটার বা মোবাইল থেকে। অবশ্য এর জন্য ইন্টারনেট লাগবে। প্রথমে `http://www.isical.ac.in/~arnabc/probstat/`-এ যাও। সেখানে দেখবে পাতার ছবি আঁকার একটা link আছে। সেটাতে ক্লিক করলেই বলে দেবে কার পরে কী করতে হবে। মোদ্দা ব্যাপারটা হল লুডোটা যতবার খেলতে চাও, (100, 200 কি 10000 যা খুশি) সেই সংখ্যাটা লেখার একটা জায়গা থাকবে, সেখানে সংখ্যাটা লিখে দিলেই কম্পিউটারটা অমনি অতবার খেলে তোমাকে ফুট্‌কির ছবিটা দেখিয়ে দেবে।

এই যে কাণ্ডটা দেখলে এটার মধ্যে আশ্চর্যের বিষয় এটাই যে যদিও আমরা প্রতি দানে ছক্কা চেলে এগোচ্ছিলাম, যেখানে কখন কী সংখ্যা আসবে তার কোনোই স্থিরতা ছিল না, অথচ সব মিলিয়ে ছবিটা দাঁড়ালো যেন ওস্তাদ শিল্পীর পাকা হাতে আঁকা! কিন্তু যদি তোমরা কয়েক বন্ধু আলাদা আলাদাভাবে একই কাজ কর, দেখবে প্রত্যেকেই ঠিক একইরকম একটা করে পাতার ছবি পাবে, যদিও ছক্কাতে সংখ্যাগুলো নিশ্চয়ই প্রতি বন্ধুর ক্ষেত্রেই একই আসে নি, কারণ সেগুলো তো এলোমেলোভাবে এসেছে। "এলোমেলো" থেকে কী করে এরকম নিখুঁত জিনিস বেরিয়ে আসে সেটাই হল statistics এবং probability-র মাথা ঘামানোর বিষয়। যাকে "এলোমেলো" বললাম, মোটামুটিভাবে তারই গালভরা নাম হল random, অর্থাৎ একই কাজ বার বার করলে বিভিন্ন সময়ে বিভিন্ন উত্তর আসে। আর যেসব ক্ষেত্রে একই কাজের ফলে সব সময়ে একই উত্তর আসে, তাকে বলে deterministic, যেমন 2 আর 3 যোগ করলে সব সময়েই 5 উত্তর আসে। সেই হিসেবে ভাবলে এখানে ছক্কার চালগুলো

Fig 4

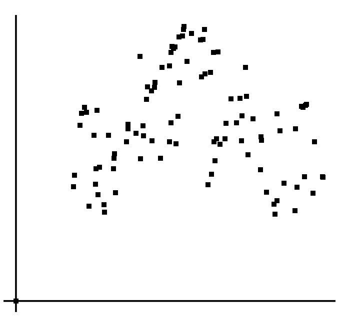


Fig 5

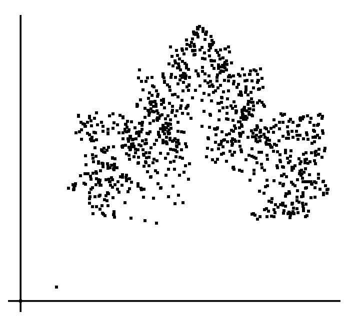
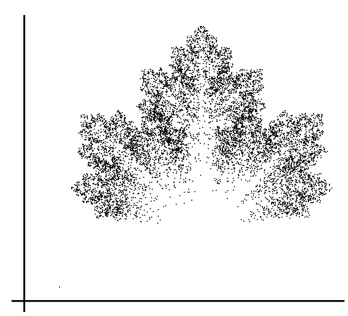


Fig 6



ছিল random, আর পাতার ছবিটা একেবারে deterministic-এর মত। একে বলে statistical regularity. ব্যাপারটা সত্যিই যেন ম্যাজিক! তুমি অনেকক্ষণ ধরে এলোমেলো কাজ করলে, অথচ সব এলোমেলো-ত্ব কেটে গিয়ে শেষমেশ একটা ভদ্র উত্তর আসতে লাগল। যেন তুমি চোখ বুঁজে তোমার মোবাইলে দশটা যা খুশি সংখ্যা টিপলে, আর সব মিলে তোমার প্রিয় বন্ধুর নম্বরটা হয়ে গেল! তাও আবার একবার নয়, বারবার!

ব্যাপারটা সত্যিই অদ্ভুত বটে, কিন্তু একটু চিন্তা করলেই বুঝবে যে এই ম্যাজিকটা তুমি আগেও দেখেছ। তোমার চোখের সামনেই এই ম্যাজিকটা আসলে অহরহ ঘটে চলেছে। আমরা আমাদের লুডোটা 10000 বার খেললে একটা পাতার ছবি পাচ্ছিলাম। যদি নতুন করে আরো 10000 বার খেলি তবে ঠিক একইরকম দেখতে আরেকটা পাতার ছবি পাব। দুটোর এমন আশ্চর্য সাদৃশ্য যে জেরক্স করলেও এত একরকম হত কিনা সন্দেহ। কিন্তু 10000-টা ছক্কার চালে প্রত্যেকবারই তো আর সবকিছু একইভাবে চলেনি নিশ্চয়ই! তার মানে গুটিটা দুই ক্ষেত্রে বিভিন্ন পথে এগিয়েছে, ফলে ফুটকিগুলোও বিভিন্ন জায়গায় বসেছে। অর্থাৎ দুটো পাতার ছবির ফুটকিগুলো সম্পূর্ণ আলাদা আলাদা জায়গায় রয়েছে। সুতরাং দুটো ছবি একেবারে এক হতেই পারে না। কিন্তু তাও কি ভীষণ কাছাকাছি! এবার একটা আম গাছের পাতাগুলোর কথা ভাবো। গাছে অনেক অনেক পাতা থাকে, সবাই একইরকম দেখতে, অথচ খুঁটিয়ে দ্যাখো--প্রত্যেকটা পাতাই কিন্তু আসলে আলাদা। আরেকটা উদাহরণ হল আমাদের আঙুলের ছাপ, চট করে দেখলে সবারই একইরকম, অথচ খুঁটিয়ে দেখলে সবারই আলাদা আলাদা! অনুমান করা অসঙ্গত নয় যে, প্রকৃতির মধ্যেও এরকম random থেকে deterministic তৈরীর খেলা অনবরত চলছে। এই ব্যাপারটাই হচ্ছে probability এবং statistics-এর মূল উপজীব্য বিষয়।

1.2 করে কী লাভ?

এবার আরেকটা উদাহরণের কথা বলি, যা থেকে বুঝবে মানুষ কেন random থেকে deterministic জিনিস বানাতে পারলে খুশি হয়। ধরো তুমি একটা কোম্পানি খুলবে। তার জন্য চাই বিদ্যুত, শ্রমিক, পণ্যপরিবহনের ব্যবস্থা। সুতরাং যদি কোনো রাজ্যে মুহূর্মুহূঃ লোডশেডিং হয়, শ্রমিকরা যখন তখন ধর্মঘট করে, আর রাস্তাঘাটের এমন করুণ অবস্থা হয় যে, পণ্যপরিবহনের কোনো স্থিরতা না থাকে, তবে বুঝতেই পারছ যে সে রাজ্যে কোম্পানির লালবাতি জ্বলতে দেরী হবে না। এটা সহজেই বোঝা যায়--তোমার কোম্পানির প্রয়োজনীয় জিনিসগুলোর (মানে বিদ্যুত, শ্রমিক ইত্যাদির) সরবরাহ যদি random হয়, তবে কোম্পানির আয়ও random হয়ে পড়বে, সেই কোম্পানি চালানো পড়তায় পোষাবে না। এবার তুমি একটা insurance কোম্পানির কথা ভাবো, যেমন ধরো জীবনবিমা। এখানে কোম্পানির মূলধন হল জনগণের জীবনমৃত্যু, যার কোনোই স্থিরতা নেই। আজকে যে লোকটা বিমা করাল সে যে একবছরের মধ্যেই গাড়িচাপা পড়ে মরবে না তার কোনো স্থিরতা নেই। কিন্তু তাতে বিমা কোম্পানি মারা পড়ছে না। বরং জীবনমৃত্যুর randomness আছে বলেই বিমা কোম্পানির জনপ্রিয়তা এবং লাভ! এবং সেই লাভটা আর পাঁচটা deterministic কোম্পানির চেয়ে কম কিছু নয়! বিমা কোম্পানির সাফল্যের পিছনে আসলে কাজ করছে probability এবং statistics-র ধারণাগুলো। একটা বাড়ির উপমা দিয়ে ভাবলে মনে করতে পারো যেন এটা এমন একটা বাড়ি যার ভিত অতিশয় নড়বড়ে, ভূমিকম্পে থরথরিয়ে কাঁপছে, অথচ তার কাঠামোর এমনি আশ্চর্য কারিকুরি যে, বাড়িটা দিব্যি বছরের পর বছর সটান দাঁড়িয়ে আছে।

1.3 রহস্যটা কোথায়?

Random জিনিসপত্র নিয়ে আমরা কি কারিকুরি করছি রহস্যটা রয়েছে সেখানেই। এর একটা নমুনা একটু আগেই দেখেছ। আমাদের অংকের লুডো খেলার নিয়মে যে বিশেষ সংখ্যাগুলো নিয়েছিলাম, সেখানেই ছিল কারিকুরিটা। যদি সংখ্যাগুলো নিতাম এইভাবে--

Fig 7

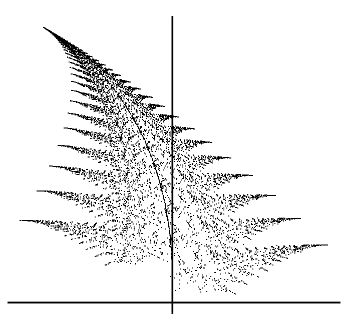




Fig 8

ছক্কায় যদি পড়ে...	তবে যাব...
1	$(\ 0,\quad 0.16y\ )$
2	$(\ 0.85x - 0.04y,\quad -0.04x + 0.85y + 1.6\ )$
3	$(\ 0.2x - 0.26y,\quad 0.23x + 0.22y + 1.6\ )$
4	$(\ -0.15x + 0.28y,\quad 0.26x + 0.24y + 0.44\ )$

তবে ছবিটা হয়ে যেত Fig 7-এর মত। আবার আরও অন্য কিছু নিলে হয়তো কোনো প্যাটার্ণই থাকত না।

এবার আরেকটা মজার খেলার কথা বলে ভূমিকার ইতি টানি। এই খেলাটার কথা পেয়েছি সুকুমার রায়ের একটা নিবন্ধ থেকে। খেলাটার জন্য খালি তিনটে ঝাঁটার কাঠি দরকার। Fig 8 দ্যাখো। দুটো কাঠিকে ভাঁজ করে V আকারের করে নাও। তৃতীয় কাঠিটাকে horizontal-ভাবে ধরো, এবং V আকৃতির কাঠি দুটোকে ওর উপর ঘোড়ায় চড়ার মত করে বসিয়ে দাও, ভিতর দিকে কিছুটা হেলিয়ে, যেমন ছবিতে দেখিয়েছি। এই কাঠিদুটোর চারটে "পা" যেন কোনো খস্‌খসে কাগজের উপর থাকে (খবরকাগজ হলেই চলবে)। এবার horizontal কাঠিটার দু প্রান্ত দু হাতে ধরে থাকো। কনুই বা হাতের অন্য কোনো অংশ যেন কোথাও ভর দিও না, এবং হাত দুটোকে সামান্য সামান্য কাঁপাও। আশ্চর্য হয়ে লক্ষ করবে যে V আকৃতির কাঠিদুটো ক্রমশঃ পরস্পরের দিকে এগিয়ে যাচ্ছে! এর কারণ হল তোমার হাতের random কম্পন। কাঠি দুটো ওইভাবে রাখার কারিকুরিতে তোমার হাতের কম্পন যতই random হোক, কাঠিগুলোর গতি সর্বদাই পরস্পরের দিকে হবে। কেন হবে, সেটা চিন্তা করে দেখতে পারো। এই বইতে আর সে প্রসঙ্গে যাব না।

1.4 অংক কষে বোঝা

1.4.1 কিছু নতুন ভাষা

এইবার তবে অংকে হাত দেওয়া যাক। Random জিনিস থেকে deterministic আচরণ তখনই আসে যখন একই random কাজ বার বার করে করা হয়। যেমন লুডোতে বার বার করে ছক্কা চালা। এইরকম random কাজকে অংকের ভাষায় বলব একটা **random experiment.** আর বারবার করে করলে প্রত্যেকবারকে বলব experiment-টার একেকটা **trial.**

Example 1: এর সবচেয়ে সহজ উদাহরণ হল একটা কয়েন টস করা। "কয়েন টস করা" ব্যাপারটা হল একটা random experiment, সেই কাজটা তুমি যতবার করবে সেগুলো হবে একেকটা trial. কোনটাকে random experiment বলব আর কোনটাকে trial বলব, সেটা নির্ধারণ করার ব্যাপারে কিছুটা স্বাধীনতা আছে। যেমন আমি চাইলে "একই কয়েনকে পরপর দুবার টস করা"-কে random experiment বলতে পারতাম, সেক্ষেত্রে পর পর দশবার টস করলে হত পাঁচটা trial. ■

Example 2: আরেকটা random experiment হল ছক্কা চালা। ■

যেহেতু ব্যাপারটা random, তাই একটা random experiment করার আগে কোনোভাবেই নিশ্চিত হয়ে বলবার জো নেই যে, কী ফলাফল হতে চলেছে। তবে কী কী ফলাফল হওয়া সম্ভব তার set-টা আমাদের জানা থাকে। এই set-টাকে বলে সেই random experiment-টার **sample space.**

Example 3: যদি random experiment-টা হয় "একবার কয়েন টস করা" তবে sample space হবে {head, tail}.

যদি random experiment-টা হত "একবার ছক্কা চালা" তবে sample space-টা হত $\{1, 2, 3, 4, 5, 6\}$. যদি random experiment-টা নিতাম "পরপর দুবার কয়েন টস করা" তবে sample space-টা হত

$$\{(\text{head}, \text{head}), \quad (\text{head}, \text{tail}), \quad (\text{tail}, \text{head}), \quad (\text{tail}, \text{tail})\}.$$

এখানে $(\text{head}, \text{tail})$ মানে হল প্রথম টসে head, আর দ্বিতীয় টসে tail পড়েছে। ■

Sample space হল একটা set. এর প্রতিটা element হল random experiment-টার একেকটা সম্ভাব্য ফলাফল, যাকে অংকের ভাষায় বলে একেকটা **outcome.** এরা যেহেতু sample space-এর element, তাই এদেরকে অনেক সময়ে **sample point**-ও বলে।

Random experiment বলতেই বারবার কয়েন টসের কথা উঠছে। এরা শুধু সহজই নয়, রীতিমত গুরুত্বপূর্ণও বটে। এর বেশ একটা গালভরা নাম আছে--Bernoulli random experiment. অবশ্য এখানে আমরা head আর tail না বলে সাধারণতঃ 1 আর 0 লিখি, কিন্তু মূল ব্যাপারটা একই। এই random experiment-এর trial-গুলোকে বলে একেকটা Bernoulli trial, যার সম্বন্ধে জানতে চেয়েছে নীচের অংকটায়।

Example 4: Write a short note on Bernoulli trials.[1] **(2009.1ap2)**

SOLUTION:

> Any random experiment with sample space $\{0, 1\}$ is called a Bernoulli random experiment. A trial of such a random experiment is called a <u>Bernoulli trial</u>. For example, if we toss a coin and denote "head" by 1 and "tail" by 0, then we get a Bernoulli trial.
>
> Similarly, any trial of a random experiment having sample space with size 2, may be considered as a Bernoulli trial.

■

Sample space-এর সঙ্গে ওতপ্রোতভাবে জড়িত হয়ে আছে আরেকটা ধারণা, তার নাম **event.** এরা হল sample space-এর একেকটা subset. একটা উদাহরণ নিই।

Example 5: আমার মোবাইলে একটা খেলা আছে, তাতে মোবাইলটা ঝাঁকালে দুটো ছক্কা চালা হয়, একটা লাল একটা সবুজ। তার মানে sample space-টা লিখতে পারি--

$$\{(i, j) \;\; : \;\; i, j = 1, 2, ..., 6\}.$$

যেমন $(1, 5)$ মানে লাল ছক্কায় 1 পড়েছে, আর সবুজ ছক্কায় 5. এটা মোটেই বিশাল মজার কিছু খেলা হল না। এটাতে একটু উত্তেজনার খোরাক যোগ করা যাক--যদি দুটো ছক্কার যোগফল হয় ≥ 10 তবে তোমার জিত, নইলে আমার জিত। যদি জিত মানে একশ টাকা পাওয়া হয়, তবে অনেকেই খেলতে লাইন দেবে। বুঝতেই পারছ এটা একধরণের জুয়ো খেলা বা যাকে ইংরাজিতে বলে gambling. বস্তুতঃ probability theory-এর গোড়াপত্তন করেছিল কিছু পাকা জুয়াড়িই, এবং জুয়োতে জেতার চেয়ে বেশী কোনো মহৎ উদ্দেশ্য তাদের ছিল না। যাই হোক, চট্ করে বলো তো sample space-এ যে ছত্রিশটা

outcome আছে, তাদের কোন কোনটার ক্ষেত্রে তোমার জিত হবে? উত্তর সহজ, সেইসব (i,j) যাদের জন্য $i+j \geq 10$ হবে, মানে--

$$\{ \quad (4,6), \quad (5,5), \quad (5,6), \quad (6,4), \quad (6,5), \quad (6,6) \quad \}$$

এইটা হল sample space-টার একটা subset, এবং একটা event-এর উদাহরণ। ■

এইখানে একটা সূক্ষ্ম সমস্যা আছে যেটার আলোচনা এই বইতে মোটেই দরকারী কিছু ছিল না, কিন্তু নেহাত নীচের অংকটার জন্য আলোচনাটা করতে হচ্ছে। Event মাত্রেই sample space-এর একেকটা subset, কিন্তু উল্টোটা সবসময়ে ঠিক নাও হতে পারে, অর্থাৎ এমন sample space-এর subset সম্ভব যারা event নয়। কেন নয়, সেই গুরু আলোচনা এখানে করা অসম্ভব। খালি এটুকু বলে রাখি যে, এই রকম গোলমেলে subset-দের দেখা সচরাচর মেলে না। বস্তুতঃ তোমাকে দশ দিন সময় দিয়ে ভাবতে বললেও তুমি এমন এক খানাও subset বার করতে পারবে না। প্রতিদিনের কাজে যেসব subset লাগে, তারা সকলেই event হয়, অতএব subset আর event-এর সুক্ষ্ম পার্থক্যটা খেয়াল না রাখলেও এই বইতে কোথাও অসুবিধা হবে না। যাই হোক, নীচের অংকটার খাতিরে বলি--যে সকল subset-রা event হয়, তাদের নিয়ে তৈরী set-টাকে নাম দেওয়া হয়েছে event space. এই event space নামটাও মোটেই প্রচলিত কিছু নয়। আন্তর্জাতিক মানের অধিকাংশ বইতেই একে বলে sigma field বা sigma algebra. যে নামেই ডাকো, এর সংজ্ঞাটা একই থাকে-- event বলতে আমরা সহজ বুদ্ধিতে যা যা স্বাভাবিক ধর্ম বুঝি তারই একটা ছোট্টো তালিকা। এটাই চেয়েছে নীচের অংকটায়।

Example 6: Write a short note on event space.[2] **(2009.1ap1)**

SOLUTION:

> Let a random experiment have sample space S. Then by the <u>event space</u> of the random experiment we mean a collection $\mathscr{F}$ of subsets of S satisfying the following conditions
>
> - $S \in \mathscr{F}$.
>   মানে পুরো sample space-টা নিজে একটা event.
> - $\forall A \in \mathscr{F} \quad A^c \in \mathscr{F}$
>   কোনো event-এর complement-ও একটা event, অর্থাৎ "head আসা" যদি একটা event হয়, তবে "tail আসা"-ও একটা event হতে বাধ্য।
> - $\forall A_1, A_2, ... \in \mathscr{F} \quad A_1 \cup A_2 \cup \cdots \in \mathscr{F}$
>
> These conditions also imply that $\mathscr{F}$ is closed under countable intersection. The elements of $\mathscr{F}$ are called events.

একটা উদাহরণ দিয়ে দিলে ভালো হবে--

> If $S = \{0,1\}$, then $\mathscr{F}_1 = \{ \ \phi, \ \{0,1\} \ \}$ is one possible event space. $\mathscr{F}_2 = \{ \ \phi, \ \{0\}, \ \{1\}, \ \{0,1\} \ \}$ is another. But $\mathscr{F}_3 = \{ \ \{0\}, \ \{1\} \ \}$ is not.

এখানে $\mathscr{F}_3$-টা একটা event space নয় কারণ এখানে $A_1 = \{0\}$ আর $A_2 = \{1\}$ নিলে তিন নম্বর শর্তটা পালিত হচ্ছে না। $A_1, A_2 \in \mathscr{F}_3$ হলেও $A_1 \cup A_2 \notin \mathscr{F}_3$.

> However, for a finite sample space S, the most common choice for event space is the power set, $\mathscr{P}(S)$.

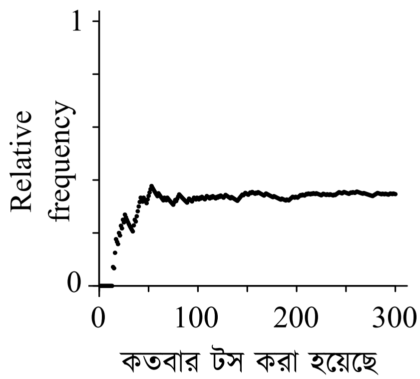


Fig 9

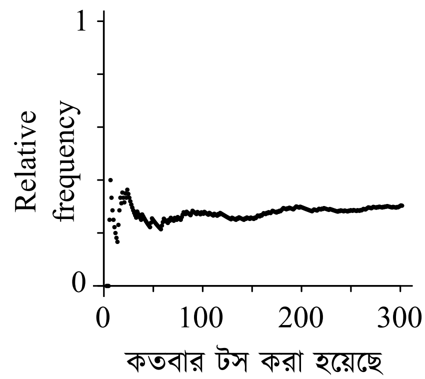


Fig 10

■

অনেকগুলো নতুন ভাষা শেখা গেল, এবার আবার statistical regularity-র প্রসঙ্গে ফিরে আসি। যেকোনো random experiment-এর যেকোনো একটা event নিলেই তার মধ্যে একটা statistical regularity-র গল্প খুঁজে পাবে। Random experiment-টা বার বার করতে থাকো। যেমন ধরো বার বার করে কয়েন টস করছ। তোমার event-টা ধরো {head}. প্রত্যেকবার দ্যাখো তোমার event-টা ঘটেছে কিনা, মানে head পড়েছে কিনা। একটা উদাহরণ নিই। মনে করো প্রথম 5-বার টসের ফলাফল হয়েছে এইরকম--

head, tail, head, head, tail.

প্রতিটা টসের পর গুণতে গুণতে যাও যে এ যাবৎ কতগুলো head দেখা গেছে। এই সংখ্যাগুলোকে বলব head পড়ার **frequency**[1]। যেমন প্রথম টসের পরে frequency-টা হল 1, যেহেতু head পড়েছে। দ্বিতীয় টসের পরেও সেই 1-ই রইল, কারণ দ্বিতীয় টসে tail এসেছে, এইরকম। একটা টেবিল হিসেবে লিখলে এরকম পাবে--

কত নম্বর টস (n)	1	2	3	4	5
Outcome	head	tail	head	head	tail
Frequency	1	1	2	3	3

এবার দেখব head পড়ার অনুপাত কত, মানে head-এর সংখ্যাগুলোকে টসের সংখ্যা দিয়ে ভাগ করব। এর ফলে যে অনুপাতগুলো পাব তাদের বলব **relative frequency**. যেমন প্রথম টসের পর head-এর frequency পেয়েছিলাম 1. তার মানে টসের সংখ্যা এখনো পর্যন্ত 1, আর head-এর সংখ্যাও 1. সুতরাং relative frequency হল $\frac{1}{1} = 1$. দ্বিতীয় টসের পর দেখতে পাচ্ছি যে, 2-টো টসে 1-টা head এসেছে, তাই relative frequency হবে $frac12$, এইরকম। টেবিলের আকারে লিখলে--

কত নম্বর টস(n)	1	2	3	4	5
Outcome	head	tail	head	head	tail
Frequency	1	1	2	3	3
Relative frequency	1	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{3}{4}$	$\frac{3}{5}$

এই উদাহরণে আমরা খালি 5 বার টস করেছিলাম। মনে করো আরো অনেকবার করলাম (যেমন ধরো 300 বার)। এবার যদি আমরা আমরা একটা গ্রাফ আঁকি, যেখানে x-axis-এ দেখাবো টসের সংখ্যা (n), আর y-axis-এ relative frequency, তবে যে ছবিটা পাবে সেটা যে একটা random ছবি তাতে নিশ্চয়ই সন্দেহ নেই? এরকম একটা গ্রাফ এঁকেছি Fig 9-এ।

[1]Frequent মানে ঘন ঘন। সেই অর্থে frequency মানে কত কত ঘন ঘন আসছে।

লক্ষ কর গ্রাফটা প্রথম দিকে যত লাফাচ্ছে, শেষের দিকে কেমন ঝিমিয়ে এসেছে? মানে ওর random ব্যাপারটা ক্রমশঃ মিলিয়ে যাচ্ছে আর তার জায়গায় একটা horizontal সরলরেখা ফুটে উঠছে। আরও মজা আছে, এবার যদি তুমি কোনো বন্ধুকে ওই কয়েনটাই আরও 300 বার টস করতে বলো (বা আঙুল ব্যথা হয়ে না থাকলে নিজেই আরও 300 বার টস কর) এবং একই কাগজের উপর ফের ওই কায়দায় গ্রাফ আঁকো তবে দেখবে যে এবার গ্রাফটা অন্যরকম হয়েছে (একটা উদাহরণ দেখিয়েছি Fig 10-এ) , কিন্তু এটাও শেষের দিকে একটা horizontal সরলরেখায় পরিণত হচ্ছে। এবং আশ্চর্যের কথা হল horizontal সরলরেখা দুটো দুই ক্ষেত্রেই একই! ছবিতে দেখে হয়তো মনে হতে পারে যে সরলরেখা দুটো সামান্য আলাদা, কিন্তু টসের সংখ্যা বাড়ালে সেই পার্থক্যটুকুও চলে যাবে। তুমি যতভাবেই কাজটা কর না কেন সবসময়েই একই জিনিস দেখবে, প্রথমে খানিকটা লাফালাফি করে গ্রাফটা ক্রমশঃ নিস্তেজ হয়ে এসে ওই বিশেষ একটা value-র দিকে এগিয়ে যাচ্ছে। এমনটা যে হবেই আমি সেটা খুবই জোর দিয়ে বলতে পারলাম কারণ এটা অংক কষে প্রমাণ করা যায়, এবং সেটা আমরা এই বইতেই শিখব। সমগ্র probability theory এবং statistics দাঁড়িয়ে আছে এই ব্যাপারটার উপরেই। প্রথমে আমরা যে লুডো থেকে পাতার ছবি পেয়েছিলাম, তার অংকটা অনেকটাই জটিল বটে, কিন্তু সেটাও আসলে এই ব্যাপারেরই রূপভেদ মাত্র।

যে ম্যাজিক সংখ্যার কাছে এসে গ্রাফটা এভাবে নিস্তেজ হয়ে মন্ত্রমুগ্ধ সাপের মত অদৃশ্য আকর্ষণে এগিয়ে যাচ্ছে তার একটা নাম আছে--**probability.** গুছিয়ে বলি। ধরো একটা event নিলাম। Event যখন বললাম, তার মানে ওর পিছনে কোনো একটা sample space আছে, আর তার মানে তারও পিছনে একটা random experiment-ও আছে[2]। এবার যদি তুমি ওই random experiment-টা বার বার করে করতে থাকো এবং event-টার relative frequency-র গ্রাফ আঁকো, তবে গ্রাফটা একটা ম্যাজিক সংখ্যার পায় লুটিয়ে পড়তে বাধ্য। সেই সংখ্যাটাকেই বলে ওই event-টার probability.

এখানে একটা কথা বলে রাখি। ওই যে বললাম বার বার করে একই random experiment করা, সেখানে কিন্তু কোনো রকম ফাঁকি চলবে না। যেমন ধরো, আঙুল ব্যথা হবার ভয়ে তুমি যদি মোটে একবার টস করে সেটাকেই 100 বার লিখে দাও, বা খালি 10 বার টস করে সেইগুলোকেই ঘুরিয়ে ফিরিয়ে বার বার করে ব্যবহার কর তবে কিন্তু গ্রাফটার কোনো গ্যারান্টি দেওয়া যাবে না, সেটা নিস্তেজ নাও হতে পারে, বা হলেও বিভিন্নক্ষেত্রে বিভিন্ন জায়গায় চলে যেতে পারে। প্রত্যেকবার random experiment-টা নতুন করে করতে হবে, যাকে অংকের ভাষায় বলে **independently** করা, মানে trial-গুলোকে independent হতে হবে। এইটা একটা গুরুত্বপূর্ণ ধারণা, যেটা আমরা এই অধ্যায়ের শেষে শীঘ্রই বিশদভাবে আলোচনা করব। আপাততঃ মূল কথাটা বুঝে রাখো, independent মানে হল একটা trial-এর কোনো প্রভাব অন্যটাতে পড়া চলবে না, ব্যস্! এইবার তবে নীচের প্রশ্নের উত্তরটা লেখার যাবতীয় রসদ মজুদ হয়েছে। যদিও গুছিয়ে লেখার কিছু প্যাঁচ আছে।

Example 7: What is meant by the term “statistical regularity”? Explain how the frequency definition of the probability of a random event is related to the concept of statistical regularity.[5] **(2009.1c)**

SOLUTION:

> Statistics deals with random experiments. The outcome of a random experiment may not be known for sure in advance, unlike deterministic mathematics (*e.g.*, $2 + 3 = 5$). Thus, we may think of randomness as ``irregularity". However, in certain situations, even outcomes of random experiments show a regularity that is similar to deterministic behaviour. This is called <u>statistical regularity</u>:
>
> If the same random experiment is repeated independently for a large number of times then certain aggregate behaviours of the outcomes often converge to a deterministic number.

এখানে "aggregate behaviour" কথাটা একটু খটোমটো, বুঝে নেওয়া যাক। পরীক্ষার নম্বরের ক্ষেত্রে আমরা aggregate বলতে বুঝি সব বিষয় মিলিয়ে মোট নম্বর। এখানেও aggregate বলতে বোঝানো হচ্ছে এখনো পর্যন্ত করা সবগুলো trial-এর

[2] ঠিক যেমন যেকোনো মানুষেরই একটা বাবা থাকে, এবং অতএব একটা ঠাকুর্দাও থাকে।

outcome মিলিয়ে তৈরী কোনো সংখ্যা। যেমন বার বার কয়েন টসের বেলায় এখনও পর্যন্ত যে কটা টস হয়েছে তাদের ভিত্তিতে head-এর relative frequency, এইরকম।

> For example, if a coin is tossed n times then the relative frequency of heads approaches a fixed number as $n \to \infty$.
> Such regular behaviours can be proved using various theorems of probability.

আমরা এরকম কিছু theorem-এর কথা পরে শিখব।

> ``Frequency definition of probability'': Suppose that we have a random experiment with sample space S. Let $A \subseteq S$ be any event. Let us repeat the random experiment n times independently. Let $X_n =$ the relative frequency of the event during these trials.
>
> Then it is seen that as $n \to \infty$ the random variable X_n approaches some fixed number. This behaviour is an example of statistical regularity. The limiting fixed number is called the probability of the event.
>
> This is called the ``frequency definition of probability''.

■

এখানে একটা কথা আছে। এই যে “definition” বললাম এটা কিন্তু একটা mathematical definition নয়। কারণ statistical regularity ব্যবহার করে mathematical definition বানাতে গেলে আগে statistical regularity ব্যাপারটা আমাদের অংক কষে প্রমাণ করা উচিত। আর সেটা করবার জন্যই probability লাগবে। তার মানে এইভাবে probability-র সংজ্ঞা দেবার জন্যই probability ব্যবহার করতে হচ্ছে! সেটা তো আর চলতে পারে না। কিন্তু চিন্তা নেই, probability-র অবশ্যই একটা জাঁদরেল mathematical definition আছে, তাকে বলে axiomatic definition. সেখানে এরকম "ডিম আগে না মুর্গী আগে" জাতীয় সমস্যা নেই। সেই সংজ্ঞাটা আমরা কালকে শিখব। কিন্তু তা বলে “frequency definition of probability”-র গুরুত্ব কিন্তু কম নয়। Statistical regularity দিয়ে চিন্তা করলে probability-র ধারণাটা যত সহজে হৃদয়ঙ্গম করা যায়, তেমনটা আর অন্য কিছুতেই হয় না।

DAY 2
Probability axioms

বুড়ো তার দুটো আঙুল দিয়ে আমায় একটুখানি টিপে-টিপে বলল, "আড়াই সের।"
আমি বললাম, 'সেকি, পটলার ওজনই তো একুশ সের, সে আমার চাইতে দেড় বছরের ছোটো।"

--হ য ব র ল (সুকুমার রায়)

গতকাল probability-র একটা গালভরা গল্প শিখেছি। গ্রাফের নিস্তেজ হয়ে আসা-টাসা নিয়ে বেশ জমাটি ব্যাপার। তবে এত কিছু না জেনেও কিন্তু probability বলতে কি বোঝায় সেটা আমরা সকলেই মোটামুটি জানি। পড়াশোনা না করলে যে পাশ করার চান্স কমে, আর মোবাইল কানে রাস্তা পার করলে যে গাড়িচাপা পড়ার চান্স বাড়ে সেটা বোঝার জন্য অংক লাগে না। সুতরাং অংক কষে আমরা যতই জম্পেশ গল্প খাড়া করি না কেন, খেয়াল রাখতে হবে যাতে আমাদের সহজ বুদ্ধির ধারণাগুলো আবার ঘেঁটে না যায়। একটা উদাহরণ নেওয়া যাক।

Example 8: একটা কোচিং ক্লাসের বিজ্ঞাপনে বলছে--

আমাদের এখানে পড়লে পাশ করার চান্স 70% আর ফার্স্ট ক্লাস পাওয়ার চান্স 80%.

সেই কোচিং ক্লাসে পড়াশোনার মান কী জানিনা, কিন্তু এই দাবীটা যে হাস্যকর সেটা নিশ্চয়ই বলে দিতে হবে না। আরে বাবা, ফার্স্ট ক্লাস পেতে হলে তো আগে পাশটুকু করতে হবে, সুতরাং ফার্স্ট ক্লাস পাওয়ার চান্স কি করে পাশ করার চান্সের চেয়ে বেশী হয়? ■

সুতরাং প্রত্যেকটা event-এর সঙ্গে যা খুশী করে একটা probability লাগিয়ে দিলেই চলবে না, সেগুলোকে কিছু নিয়ম মেনে চলতে হবে। Kolmogorov নামে একজন রাশিয়ান গণিতজ্ঞ এইরকম নিয়মের একটা তালিকা বানিয়েছিলেন। তালিকার শর্তগুলোকে বলে probability-র একেকটা axiom (স্বতঃসিদ্ধ)। এই axiom-গুলো ব্যবহার করেই অংকের ভাষায় probability-র সংজ্ঞা দেওয়া হয়। এই সংজ্ঞাকে বলে axiomatic definition of probability.

Example 9: State precisely the axiomatic definition of probability. [2]

(2005.1a)
SOLUTION:

> DEFINITION: **Axiomatic definition of probability**
>
> Let S be a nonempty set (called the sample space). Let $\mathscr{F}$ be the set of all events (subsets of S). Then a function $P : \mathscr{F} \to \mathbb{R}$ is called a probability if it satisfies the following conditions:
>
> 1. $\forall A \in \mathscr{F} \quad P(A) \geq 0$.
> 2. $P(S) = 1$.
> 3. For any countable collection of pairwise disjoint events $A_1, A_2, ... \in \mathscr{F}$
> $$P(\cup_n A_n) = \sum_n P(A_n).$$

■

Exercise 1: State the axioms of probability. **(2010.1a)**
HINT: আগের অংকটাই। ■

Axiom-গুলো দেখতে যতই খটোমটো লাগুক, প্রত্যেকটাই কিন্তু সহজ বুদ্ধি থেকেই আসে। যেমন ধরো প্রথম axiom-টা, যেটা বলছে যে probability সবসময়ে ≥ 0 হয়। সেটা তো আমরা এমনিতেই জানি, কেউ কি কখনও বলে যে পরীক্ষায় পাশ করার চান্স -50%? একইভাবে দ্বিতীয় axiom-টা দ্যাখো। যেহেতু S হল পুরো sample space-টা, তাই যাই ঘটুক না কেন outcome-টা ওর মধ্যে থাকতে বাধ্য। ঠিক যেমন তোমার যদি একটা কঠিন অপারেশন হয়, তবে অপারেশনের পর হয় তুমি বেঁচে থাকবে আর নয় তো মরে যাবে, এর বাইরে তো আর কিছু হতে পারে না! সুতরাং এখানে $S = \{$ বাঁচা, মরা $\}$. এবং $P(S) = 1$. কোনো event-এর probability=1 মানে সেটা ঘটবেই। তিন নম্বর axiom-টা একটু বেশী জটিল দেখাচ্ছে, কিন্তু সেটাও সহজ বুদ্ধিরই জিনিস। ধরো তুমি একটা competitive পরীক্ষায় বসবে। সেখানে analysis আর algebra দুটো থেকেই সমান সমান প্রশ্ন আসে। তুমি analysis-এ কাঁচা, তাই analysis-এর প্রশ্নগুলো পারার চান্স তোমার পক্ষে কম (ধরো 20%)। কিন্তু algebra-য় তোমার জোর বেশী, সেখানকার প্রশ্নগুলোর ঠিক উত্তরের চান্স হল মনে

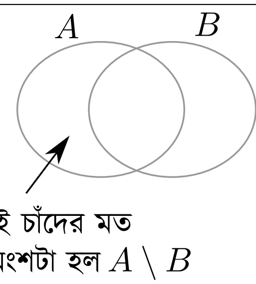


Fig 11

করো 60%. সুতরাং তুমি নিশ্চয়ই এভাবে ভাববে--algebra থেকে মোটামুটি 60% পাচ্ছি আর analysis থেকে 20% সুতরাং মোটামুটি $60\% + 20\% = 80\%$ নম্বর তোলার চান্স আছে।
Axiom-গুলোর যে ব্যাপারটা সহজ বুদ্ধিতে ধরা পড়ে না সেটা হল এর উল্টো দিকটা-- probability নিয়ে যা যা জিনিস তুমি সহজ বুদ্ধিতে (বা জটিল বুদ্ধিতেও) ভাবতে পারো, সব কিছুই প্রমাণ করা যায় মাত্র এই তিনটে axiom দিয়েই! Axiom-গুলোর বাহাদুরিটা এইখানেই! তারই খানিক পরিচয় পাব এইবার।

2.1 Axiom-গুলো দিয়ে কী করা যায়?

সহজ বুদ্ধিতে probability-র যা যা গুণ থাকা কাম্য সবগুলোই axiom-গুলো থেকে প্রমাণ করা যায়। সেসব প্রমাণের লম্বা ফিরিস্তি নিয়েই অনেকক্ষণ কাটিয়ে দেওয়া সম্ভব। তবে তাতে জ্ঞানবৃদ্ধির চাইতে বিরক্তিবৃদ্ধি হবে বেশী। তাই আমরা খালি তিন ধরণের জিনিসের স্বাদ দেব-- প্রথমে একেবারে সহজ, তারপর মাঝারি, সবশেষে একটু কঠিন।

2.1.1 সহজ

Example 10: ধরো A আর B দুটো event, যেখানে $B \subseteq A$. তবে দেখাও যে $P(B) \leq P(A)$ হতে বাধ্য।

SOLUTION: এইটা আসলে সেই কোচিং ক্লাসের গল্পটাই। A হল "পাশ করা" আর B হল "ফার্স্ট ক্লাস পাওয়া"।

> Let $C = A \setminus B$.

মানে A থেকে B বাদ দিলে যা পড়ে থাকে (Fig 11)।

> Then $B \cap C = \phi$ and $B \cup C = A$.
> So
> $$\begin{aligned} P(A) = P(B \cup C) &= P(B) + P(C) \quad [\text{ by axiom 3 } \because B \cap C = \phi] \\ &\geq P(B) \quad [\because P(C) \geq 0 \text{ by axiom 1}] \end{aligned}$$

■

Example 11: Show that for any event A, we have $P(A^c) = 1 - P(A)$.

SOLUTION:

> We know that $A \cup A^c = S$, the whole sample space.
> Also $A \cap A^c = \phi$.
> By axiom 2, we know $P(S) = 1$.

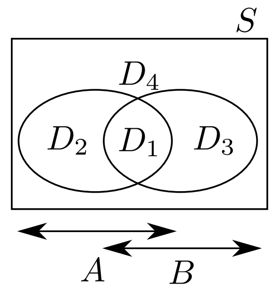


Fig 12

By axiom 3, we have $P(S) = P(A) + P(A^c)$.
Combining, $P(A) + P(A^c) = 1$.
So $P(A^c) = 1 - P(A)$, as required.

■

Example 12: Show that for any two events A and B belonging to the event space S,

$$P(\bar{A}\bar{B}) = P(\bar{A}) - P(B) + P(AB),$$

where $\bar{A} = S - A$.[3]
(2005.1a)
SOLUTION: এই অংকে $\bar{A} < \bar{B}$ মানে হল যথাক্রমে A^c, B^c. আমরা অবশ্য A^c, B^c-ই লিখব এ বইতে। আরেকটা notation ব্যবহার করেছে এখানে, AB মানে হল $A \cap B$. একইভাবে $\bar{A}\bar{B}$ মানে $A^c \cap B^c$.
এরকম অংকে প্রথমেই Fig 12-এর মত একটা ছবি এঁকে নিলে সুবিধা হয়, যাকে বলে Venn diagram. আমাদের set দুটো হল A আর B, যাদেরকে ডিমের মত দেখিয়েছি। লক্ষ কর যে, পুরো ছবিটা যেন চারটে টুকরো পাশাপাশি বসিয়ে তৈরী। টুকরোগুলোকে $D_1, ..., D_4$ নাম দিয়েছি। যেমন D_1 হল $A \cap B$, এইরকম। ভাবতে পারো যেন Venn diagram-এর লাইনগুলো বরাবর কাঁচি চালালে যে টুকরোগুলো পাবে তারাই হল D_i. এখানে A, B-র বদলে এই D_i-দের নিয়ে কাজ করার সুবিধা হল এই যে, D_i-রা সবাই disjoint. তাই axiom 3-টা নিশ্চিন্তে লাগানো যাবে। প্রথমে D_i-দের পরিচয়লিপি লিখে নেওয়া ভালো--

Let

$$D_1 = A \cap B, \quad D_2 = A \cap B^c, \quad D_3 = A^c \cap B, \quad D_4 = A^c \cap B^c.$$

Clearly, these are all pairwise disjoint.

যেহেতু সবকিছুই এই D_i-দের দিয়ে তৈরী, তাই যাই প্রমাণ করতে দিক না কেন এদের দিয়ে সেটা করা যাবে। আমাদের অংকে আমরা $A^c, B, A \cap B$ আর $A^c \cap B^c$ এই চারটে set নিয়ে কাজ করছি। এদেরকে D_i দিয়ে লিখে নেওয়া যাক।

$$\begin{aligned} A^c \cap B^c &= D_4, \\ A^c &= D_3 \cup D_4, \\ B &= D_1 \cup D_3, \\ A \cap B &= D_1. \end{aligned}$$

ব্যস্, এবার বাঁদিক আর ডানদিক দুদিককেই D_i দিয়ে লিখে ফেললে মিলে যেতে বাধ্য!

The L.H.S. is $P(A^c \cap B^c) = P(D_4)$.
The R.H.S. is

$$\begin{aligned} & P(A^c) - P(B) + P(A \cap B) \\ = & P(D_3 \cup D_4) - P(D_1 \cup D_3) + P(D_1) \\ = & \big[P(D_3) + P(D_4)\big] - \big[P(D_1) + P(D_3)\big] + P(D_1) \quad [\because D_i\text{'s pairwise disjoint}] \\ = & P(D_4), \end{aligned}$$

same as the L.H.S, as required.

■

এবার একই কায়দায় নীচের দুটো অংক করো দেখি! ওরা আসলে একই অংক।

Exercise 2: For any two events A and B, show that

$$P(A) = P(A \cap B) + P(A \cap B^c).$$

■

Exercise 3: Establish the relation

$$P(A' \cap B) = P(B) - P(A \cap B).$$

[3] **(2014)** ■

এই একই কায়দায় একগুচ্ছ দরকারী জিনিস প্রমাণ করা যায়, যাদেরকে বলে **inclusion-exclusion principle.** একটু নমুনা দেখা যাক--

Example 13: For any two events A, B show that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

SOLUTION: এখানেও আমরা সেই Fig 12 নিয়েই কাজ করছি। বস্তুতঃ দুটো event নিয়ে এরকম যেকোনো অংকই এই ছবিটা দিয়ে করা যায়।

Let

$$\begin{aligned} D_1 &= A \cap B, \\ D_2 &= A \cap B^c, \\ D_3 &= A^c \cap B, \\ D_4 &= A^c \cap B^c. \end{aligned}$$

Then D_i's are all pairwise disjoint.

Also $A = D_1 \cup D_2$ and $B = D_1 \cup D_3$ and $A \cup B = D_1 \cup D_2 \cup D_3$.
So the L.H.S. is $P(A \cup B) = P(D_1 \cup D_2 \cup D_3) = P(D_1) + P(D_2) + P(D_3)$, by axiom 3, since the events are pairwise disjoint.

The R.H.S. is

$$\begin{aligned} & P(A) + P(B) - P(A \cap B) \\ = \; & P(D_1 \cup D_2) + P(D_1 \cup D_3) - P(D_1) \\ = \; & P(D_1) + P(D_2) + P(D_1) + P(D_3) - P(D_1) \quad [\because D_i\text{'s pairwise disjoint}] \\ = \; & P(D_1) + P(D_2) + P(D_3), \end{aligned}$$

same as the L.H.S., as required.

■

এই যেটা প্রমাণ করলাম, সেটাকে কেন inclusion-exclusion principle বলে, সেটা বোঝা কঠিন নয়। আমরা $P(A \cup B)$ বার করতে চাইছি। $A \cup B$ তৈরী হয়েছে A আর B দুজনকে নিয়েই। সুতরাং প্রথমেই $P(A) + P(B)$-কে হিসেবে নিলাম (মানে include করলাম)। কিন্তু তাতে আবার $A \cap B$ অংশটা দুবার হিসেবে এসে গেল, তাই ওটাকে একবার বাদ দিলাম (মানে exclude করলাম)। সেই জন্যই ওই $-P(A \cap B)$-টা এসেছে।
একই কাজ করা যায় তিন বা তারও বেশী সংখ্যক event নিয়েও। যেমন তিনটে event নিলে কী হবে সেটা নীচের অংকে দিয়েছি।

Exercise 4: For any three events A, B and C, show that

$$P(A \cup B \cup C) = \underbrace{P(A) + P(B) + P(C)}_{\text{inclusion}} \underbrace{- P(A \cap B) - P(B \cap C) - P(C \cap A)}_{\text{exclusion}} + \underbrace{P(A \cap B \cap C)}_{\text{inclusion}}.$$

HINT: কী করে এগোতে হবে, সেটা একটু ধরিয়ে দিই। Fig 13 দ্যাখো। এখানে set-গুলোর সীমানা বরাবর কাঁচি চালালে পাবে 8-টা টুকরো, যাদেরকে $D_1, ..., D_8$ নামে দেখিয়েছি। উদাহরণস্বরূপ, $D_2 = A^c \cap B \cap C$. এবার বাঁ আর ডানদিককে D_i-দের দিয়ে লিখে ফেললেই মিলে যাবে।
এখানে একটা ছোটো কায়দা বলি যা দিয়ে তুমি চট্ করে বার করতে পারবে কোন D_i-টা কী হবে। ধরো D_5-এর জন্য বার করব। প্রথমে $A \cap B \cap C$ লিখে শুরু করো। এবার Fig 13 থেকে দেখে নাও, D_5-টা A, B, C কার কার মধ্যে আছে। এটা আছে B-র মধ্যে, A-র বাইরে, C-এর বাইরে। যেগুলোর বাইরে তাদের complement নিতে হবে। ফলে পাবে

$$D_5 = A^c \cap B \cap C^c.$$

■

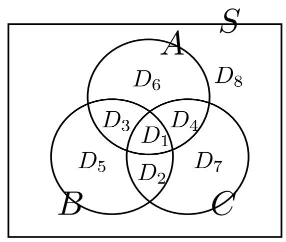


Fig 13

2.1.2 মাঝারি

এবার আমরা কিছু অংক দেখব যেগুলো আগের থেকে সামান্য বেশী কঠিন। আগের অংকগুলোয় দুটো বা তিনটে event থাকছিল। এবার থাকবে n-খানা করে।

Example 14: For any n events $A_1, A_2, ..., A_n$, show that

$$P(A_1 \cdot A_2 \cdots A_n) \geq P(A_1) + P(A_2) + \cdots + P(A_n) - (n-1).$$

Hence deduce that

$$P(A_1 + A_2 + \cdots + A_n) \leq P(A_1) + P(A_2) + \cdots + P(A_n).$$

[2+3]

(2012,2010,2007)

SOLUTION: এখানে $A_1 \cdot A_2 \cdots A_n$ মানে $A_1 \cap A_2 \cap \cdots \cap A_n$. আর $A_1 + A_2 + \cdots + A_n$ মানে $A_1 \cup A_2 \cup \cdots \cup A_n$. আমরা এখানে induction লাগাব। মনে আছে নিশ্চয়ই যে induction-এর তিনটে অংশ হয়--basis, hypothesis আর step.

First part: (Induction on n):

Basis: For $n = 1$ the LHS $= P(A_1) =$ RHS.

For $n = 2$ $P(A_1 \cup A_2) \leq 1$ and $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$.

So

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1) + P(A_2) - P(A_1 \cup A_2) \\ &\geq P(A_1) + P(A_2) - 1 \quad [\because P(A_1 \cup A_2) \leq 1] \\ &= P(A_1) + P(A_2) - (n-1), \end{aligned}$$

as required.

Hypothesis: Let us assume the result for $n = 1, ..., k$ for some $k \geq 2$.

Step: Shall prove for $n = k + 1$.

So

$$P(\underbrace{A_1 \cap \cdots \cap A_k}_{B} \cap A_{k+1}) \geq P(B) + P(A_{k+1}) - 1 \quad [\text{ induction hyp. for } 2 \text{ events}]$$

$$\geq \big[P(A_1) + \cdots + P(A_k) - (k-1)\big] + P(A_{k+1}) - 1$$

$$[\text{ induction hyp. for } k \text{ events}]$$

$$\geq P(A_1) + \cdots + P(A_k) + P(A_k) - (n-1) \quad [\because n = k+1.],$$

as required, completing the proof by the principle of mathematical induction.

এবার দ্বিতীয় অংশ। এখানে ব্যবহার করব de Morgan's law, যেটা বলে $A_1, ..., A_n$ কিছু set হলে $(A_1 \cup \cdots \cup A_n)^c = A_1^c \cap \cdots \cap A_n^c$ হয়।

Second part: Let $A_1, ..., A_n$ be any events. We apply the first part to $A_1^c, ..., A_n^c$ to get

$$P(A_1^c \cap \cdots \cap A_n^c) \geq P(A_1^c) + P(A_2^c) + \cdots + P(A_n^c) - (n-1).$$

By de Morgan's laws we know that

$$A_1^c \cap \cdots \cap A_n^c = (A_1 \cup \cdots \cup A_n)^c.$$

Hence

$$P\big((A_1 \cup \cdots \cup A_n)^c\big) \geq P(A_1^c) + P(A_2^c) + \cdots + P(A_n^c) - (n-1).$$

Now, for any event B we have $P(B^c) = 1 - P(B)$. So

$$1 - P(A_1 \cup \cdots \cup A_n) \geq [1 - P(A_1)] + [1 - P(A_2)] + \cdots + [1 - P(A_n)] - (n-1),$$

or

$$P(A_1 \cup \cdots \cup A_n) \leq P(A_1) + P(A_2) + \cdots + P(A_n),$$

as required.

■

Exercise 5: Prove that $P(A_1 + \cdots + A_n) \leq P(A_1) + \cdots + P(A_n)$, where $A_1, ..., A_n$ are n random events.[5]
(2005.1b)
HINT:
এই অংকে random event বলে একটা কথা ব্যবহার করা হয়েছে, যার কোনো মানে হয় না। কোনো event কখনও random হয় না, ওরা একটা random experiment-এর sample space-এর একেকটা subset বটে, কিন্তু ওরা নিজেরা মোটেই random নয় (অর্থাৎ ভাগ্যের উপর নির্ভর করে বদলে বদলে যায় না)। যেমন একটা ছক্কা চালা হলে sample

space হল $\{1,2,3,4,5,6\}$. এখানে "odd সংখ্যা আসা" event-টা হল $\{1,3,5\}$. এটা চমৎকার একটা deterministic set, কোনো randomness নেই। যাই হোক, এখানে ওই random শব্দটা অগ্রাহ্য করতে হবে, এবং সেক্ষেত্রে এটা আগের অংকের দ্বিতীয় অংশে পরিণত হবে। সুতরাং তুমি চাইলে সেভাবে এগোতে পারো। কিন্তু তাহলে সমাধানটা বড্ড লম্বা হয়ে যাবে, কারণ দ্বিতীয় অংশটার জন্য আগে তোমাকে তাহলে প্রথম অংশটা প্রমাণ করে নিতে হবে। এখানে বুদ্ধিমানের কাজ হবে সরাসরি প্রমাণ করা induction ব্যবহার করে। সেটা তোমার উপরই ছেড়ে দিচ্ছি, খালি induction step-এর মূল যুক্তিটা বলে দিই-- ধরো $k=1,2,...,n$-এর জন্য $P(A_1\cup\cdots\cup A_k)\leq P(A_1)+\cdots+P(A_k)$ দেখানো হয়েছে। এবার দেখাতে হবে $k=n+1$-এর জন্য। তাহলে $P(\underbrace{A_1\cup\cdots\cup A_n}_{B}\cup A_{n+1})\leq P(B)+P(A_{n+1})$ হবে। আর $P(B)\leq P(A_1)+\cdots+P(A_k)$ তো দেখানো হয়েইছে। সুতরাং হয়ে গেল! অবশ্য খালি "হয়ে গেল" বলে ছেড়ে দিলেই চলবে না, খুঁটিনাটিগুলো লিখতে হবে। সেগুলো আগের অংকের প্রথম অংশের অনুকরণে লেখার ভার তোমার উপর রইল। ■

2.1.3 কঠিন

আমরা probability axiom-গুলোর বিভিন্ন প্রয়োগ দেখছিলাম। প্রথমে কিছু সহজ প্রয়োগ দেখেছি যেখানে আমরা দুটো বা তিনটে event নিয়ে কাজ করছিলাম, তারপর এল মাঝারি কিছু অংক যেখানে n-খানা event নিয়ে কাজ হল। এবার আমরা countably infinite সংখ্যক event নিয়ে কাজ করব, মানে $A_1,A_2,A_3,...$।

Example 15: Write a short note on increasing sequence of events.[2] **(2009.1a)**

SOLUTION:

> Let a random experiment have sample space S and event space $\mathscr{F}$. Then by an increasing sequence of events we mean a sequence $A_1,A_2,...\in\mathscr{F}$ such that
>
> $$A_1\subseteq A_2\subseteq\cdots.$$
>
> For example, if the random experiment is tossing a coin 10 times, and A_n is the event "there is at least one head in the first n tosses" then $A_1\subseteq A_2\subseteq\cdots\subseteq A_{10}$.

এটা কেন হল বুঝলে তো? A_1 যদি ঘটে থাকে, তবে প্রথম টসেই head এসেছে, সুতরাং দ্বিতীয় টসে যাই আসুক না কেন প্রথম দুটো টসে অন্ততঃ একটা head তো হয়েই গেল! তাই A_2 হতে বাধ্য। সেই কারণেই $A_1\subseteq A_2$ হয়েছে। একইভাবে $A_2\subseteq A_3$ ইত্যাদিও হবে।

> The sequence may be finite or infinite. If it is infinite, then its limit is defined as $\cup_n A_n$.
> If P is a probability then clearly
>
> $$P(A_1)\leq P(A_2)\leq\cdots.$$
>
> It can be proved that for an infinite sequence of increasing events
>
> $$\lim_{n\to\infty}P(A_n)=P(\lim_{n\to\infty}A_n).$$

■

এই শেষ লাইনের প্রমাণটা রয়েছে নীচের অংকটায়। সেখানে অবশ্য কাজ করতে বলেছে monotone sequence of events নিয়ে। সেটার মানে কী নিশ্চয়ই আন্দাজ করতে পারছ। Increasing মানে যদি $A_1 \subseteq A_2 \subseteq \cdots$ হয়, তবে decreasing মানে যে $A_1 \supseteq A_2 \supseteq \cdots$ হবে বোঝাই যায়। এদের দুজনকে একসঙ্গে বলে monotone sequence, অর্থাৎ যেসব sequence হয় increasing নয়তো decreasing, তারা। একটা increasing sequence-এর উদাহরণ তো আগের অংকেই দেখেছি, কয়েন টসের উদাহরণে। সেটা ছিল increasing sequence. ওটাকেই একটু এদিক ওদিক করে একটা decreasing sequence-এর উদাহরণ বানানো যায় এইভাবে-- পর পর টস করে যাওয়া হচ্ছে, B_n হল "প্রথম n-টা টসে একবারও head পড়ে নি"। তাহলে

$$B_1 \supseteq B_2 \supseteq B_3 \supseteq \cdots$$

হবে। এখানে $B_1 \supseteq B_2$ কেন হচ্ছে, বুঝে নেওয়া যাক। যদি B_2 হয়, তার মানে প্রথম দুটো টসের একটাতেও head পড়ে নি। সুতরাং প্রথম টসেও head পড়ে নি, তাই B_1-ও হতে বাধ্য। একইভাবে $B_2 \supseteq B_3$ ইত্যাদিও হবে।
এবার দুটো উদাহরণ দেখি যারা monotone নয়। ধরো আমাদের sample space হল $\{1, 2, 3, ...\}$. যদি sequence-টা এরকম নিই--

$$A_1 = \{1\}, \quad A_2 = \{2\}, ...$$

তবে monotone হবে না, কারণ এখানে $A_1 \subseteq A_2$ নয়, আবার $A_2 \subseteq A_1$-ও নয়। আবার যদি

$$A_1 = \{1\}, \quad A_2 = \{1, 2\}, A_3 = \{2\}, A_4 = \{2, 3\}, ...$$

নিতাম, তবেও হত না, কারণ এখানে A_1-এর চেয়ে A_2 বড়, সুতরাং increase করছিল, কিন্তু তারপরেই আবার A_3-টা ছোটো হয়ে গেল। এরকম কখনো increase কখনও decrease করলে তাকে monotone বলা যাবে না।

Example 16: If $\{A_n\}$ is a monotone sequence of events then show that

$$P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n).$$

[5] **(2012,2008,2006)**
SOLUTION: এই অংকটা যা প্রমাণ করতে দিয়েছে সেটা probability-র দুনিয়ায় একটা বেশ গুরুত্বপূর্ণ জিনিস। প্রথমে আমাদের পরিচিত একটা ব্যাপারের সঙ্গে সাদৃশ্য লক্ষ কর। যদি বলতাম যে $f : \mathbb{R} \to \mathbb{R}$ এমন একটা function, যাতে $a_n \to a$ হলেই $f(a_n) \to f(a)$ হয়, তবে f-এর সম্বন্ধে কী বলা যায়? উত্তর হল--f তাহলে continuous হতে বাধ্য। সেইভাবে দেখলে এখানে যেটা দেখাতে বলেছে তার মানে করা যায়--probability একটা continuous function.
আমরা প্রমাণটা দুই ভাগে করব, প্রথমে increasing-এর জন্য, তারপর decreasing-এর জন্য।

> Case I: Increasing sequence:
>
> Let $A_1 \subseteq A_2 \subseteq \cdots$ be an increasing sequence of events.

Fig 14-এর দিকে চোখ রাখো। ছবিতে খালি কয়েকটা A_n দেখিয়েছি। কল্পনা করে নাও এরকম infinitely-সংখ্যক আছে। যদি set-গুলোর সীমানা বরাবর কাঁচি চালাই তবে টুকরোগুলো কীরকম হবে সেটা Fig 15-এ দেখিয়েছি। প্রথম টুকরোটা A_1 নিজেই, একটা চাকতির মত। তারপর থেকে পরপর পেঁয়াজের খোসার মত রিঙ বেরোবে। এই টুকরোগুলোর নাম দিলাম $B_1, B_2, ...$ ইত্যাদি, অর্থাৎ--

> Define $B_1 = A_1$ and, for $n \geq 2$, $B_n = A_n \setminus A_{n-1}$.

$\{A_n\}_n$-এর বদলে $\{B_n\}_n$-দের নিয়ে কাজ করার সুবিধা কী সে তো বুঝতেই পারছ। এক--

> Clearly, B_n's are pairwise disjoint.

তাই ওদের উপর axiom 3 লাগানো যাবে। আর দুই, ওদের জুড়ে জুড়ে A_n-দের এবং $\lim_{n\to\infty} A_n$-দের লেখা যায়, যাদের দিয়ে আমাদের অংকটার L.H.S. এবং R.H.S. তৈরী--

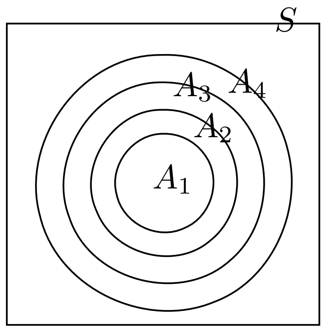


Fig 14

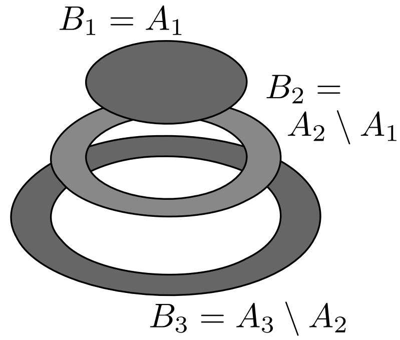


Fig 15

> Also
> $$\forall n \in \mathbb{N} \quad A_n = \cup_1^n B_k.$$
> and
> $$\lim_{n\to\infty} A_n = \cup_1^\infty A_k = \cup_1^\infty B_k.$$
> So
> $$\begin{aligned} P(\lim_{n\to\infty} A_n) &= P\left(\cup_1^\infty B_k\right) \\ &= \sum_1^\infty P(B_k) \quad [\text{ by axiom } 3, \because B_k\text{'s pairwise disjoint}] \\ &= \lim_{n\to\infty} \sum_1^n P(B_k) \quad [\text{ by defn of infinite series}] \\ &= \lim_{n\to\infty} P(\cup_1^n B_k) \quad [\text{ by axiom } 3, \because B_k\text{'s pairwise disjoint}] \\ &= \lim_{n\to\infty} P(A_n), \end{aligned}$$
> as required.

এবার nonincreasing কেসটা করব। নতুন করে কিছুই করতে হবে না, স্রেফ complement নিয়ে আগের কেসটা লাগিয়ে দিলেই হবে--

> <u>Case II: Decreasing sequence:</u>
> Let $C_1 \supseteq C_2 \supseteq \cdots$.
> Define $A_n = C_n^c$.
> Then $P(A_n) = 1 - P(C_n)$.
> Also $A_1 \subseteq A_2 \subseteq \cdots$.
> So, by <u>case I</u>,
> $$\lim_{n\to\infty} P(A_n) = P(\cup A_n). \qquad (*)$$

আমাদের প্রথম কাজ হল $\lim_{n\to\infty} P(C_n)$-কে $\lim_{n\to\infty} P(A_n)$ দিয়ে লেখা--

> Hence
>
> $$\begin{aligned}\lim_{n\to\infty} P(C_n) &= \lim_{n\to\infty}[1-P(A_n)] \\ &= 1-\lim_{n\to\infty} P(A_n) \\ &\qquad \text{ব্যস্, প্রথম উদ্দেশ্য সফল, এবার (*) লাগাব} \\ &= 1-P(\cup_1^\infty A_n) \quad [\text{ by } (*)] \\ &\qquad \text{এবার আবার } A_n\text{-দের থেকে } C_n\text{-এ ফিরে যাব} \\ &= P\left((\cup_1^\infty A_n)^c\right. \\ &= P(\cap_1^\infty A_n^c) \\ &= P(\cap C_n) \\ &= P(\lim_{n\to\infty} C_n),\end{aligned}$$
>
> as required.

■

Example 17: A coin is tossed until there are either two consecutive heads or two consecutive tails or the number of tosses becomes 5. Describe the sample space and find the probability of each sample point if every sequence n tosses has probability 2^{-n}.[5]
(2010.1d)

SOLUTION: একটা লম্বা গল্প বলেছে, মানেটা বুঝে নেওয়া যাক কয়েকটা উদাহরণ দিয়ে। একটা কয়েন আছে, সেটাকে বারবার টস করা হচ্ছে, যতক্ষণ না পরপর দুটো একই জিনিস আসে (মানে পরপর দুটো head, বা পরপর দুটো tail)। যদি কিছুতেই এরকম না হয় (মানে head আর tail ক্রমাগতই একটা অন্তর একটা আসতে থাকে) তবে 5 বার টস করার পর আমরা হাল ছেড়ে দেব। প্রথম প্রশ্ন হল এরকমভাবে চললে কী কী হতে পারে। কয়েকটা উদাহরণ কল্পনা করা যাক। ধরো প্রথম দুটো টসেই head এল। ব্যস্, আমরা আর টস করব না। তার মানে একটা outcome পেলাম HH. যদি প্রথম টসে head, দ্বিতীয় টসে tail আসত তবে কিন্তু আরও টস করতাম। যদি তৃতীয় টসে tail পেতাম, অমনি ফের থেমে যেতাম (এর ফলে outcome-টা হত HTT), যদি তৃতীয় টসে head আসত তবে ফের টস করতে হত। আচ্ছা বল তো $THTTH$ কি একটা outcome হতে পারে? না পারে না, কারণ তিন আর চার নম্বর টসে যেই TT পেয়ে গেছি, অমনি থেমে যাওয়ার কথা ছিল। এই সব উদাহরণ মাথায় রেখে এবার সমাধানটা পড়--

> The sample space consists of a sequence of H's and T's satisfying the following conditions:
>
> - the sequence has length at least 2 and at most 5.
> - for sequences with length < 5, the last two outcomes are the same, and the earlier outcomes alternate.
> - for sequences with length 5, the outcomes up to the 4-th position alternate.

Thus the sample space is

$\{HH, TT, HTT, THH, HTHH, THTT, HTHTT, HTHTH, THTHH, THTHT\}$.

The probabilities of a sequence of length n is 2^{-n}. So the probabilities are

Sample point	HH	TT	HTT	THH	$HTHH$	$THTT$	$HTHTT$
Probability	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{32}$

Sample point	$HTHTH$	$THTHH$	$THTHT$
Probability	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{32}$

টেবিলটা লম্বা হয়ে যাচ্ছিল বলে দুই ভাগে ভেঙে দেখিয়েছি। ■

DAY 3
Equally likely outcomes (part 1)

আমরা গতকাল axiomatic definition of probability শিখেছি। এর জন্য তিনটে উপাদান লাগে--

- এক, একটা nonempty set, যাকে বলে sample space. এই বইতে সাধারণতঃ আমরা sample space-টাকে S নামে ডাকছি।
- দুই, S-এর এক গুচ্ছ subset, যাদেরকে বলে একেকটা event. এরা এই শর্তগুলো পালন করে--
  - S নিজে একটা event
  - যদি A একটা event হয়, তবে A^c-ও একটা event হবে।
  - যদি $A_1, A_2, ...$ ইত্যাদিরা সবাই event হয়, তবে এদের union এবং intersection-রাও event হবে।
- আর তৃতীয় উপাদান হল একটা probability function, যেটা axiom-গুলো মেনে চলবে।

এবার আমরা এর দুটো বিশেষ কেস দেখব, যাদের বলা যেতে পারে সবচেয়ে সহজ দুটো কেস। এদের দুজনেরই মূল ব্যাপার হল সব outcome-এরই সমান সম্ভাবনা। পার্থক্য খালি একটাই, প্রথম ক্ষেত্রে sample space-টা হবে finite, আর দ্বিতীয় ক্ষেত্রে infinite.

3.1 Finite sample space

এখানে কী নিয়ে কাজ করব বলে রাখি, মানে axiomatic definition-এর তিনটে উপাদান কী কী নেব--

- Sample space-টা নেব যেকোনো nonempty finite set. যেমন ধরো $S = \{1, 2, 3\}$, বা $S = \{$ সাদা, কালো, লাল, নীল$\}$, কিংবা $S = \{$ হোমস, ফেলুদা, ব্যোমকেশ$\}$.
- S-এর প্রতিটা subset-ই একেকটা event. যেমন ধরো $S = \{1, 2\}$ হলে তার চারটে subset সম্ভব, এরা সকলেই event– $\mathscr{P}(S) = \{ \quad \phi, \quad \{1\}, \quad \{2\}, \quad \{1, 2\} \quad \}$.

- আমরা probability নেব এইভাবে--যদি কোনো event থাকে $A \subseteq S$, তবে তার probability হবে

$$P(A) = \frac{|A|}{|S|}.$$

  এখানে $|A|$ মানে A-র মধ্যে কতগুলো element আছে, আর $|S|$ মানে S-এর মধ্যে কতগুলো element আছে। যেহেতু S হল finite, তাই A-ও finite. সুতরাং $|A|/|S|$ বার করতে কোনো অসুবিধা নেই (যেহেতু $S \neq \phi$, তাই $|S| \neq 0$)।

একটা ছোট্টো উদাহরণ করে সড়গড় হয়ে নিই।

Example 18: Sample space নাও $S = \{1, 2, ..., 10\}$. যদি সবগুলো outcome-ই equally likely হয়, তবে $P(\{1, 4, 5\})$ কত হবে? $P(\{1\})$ বেশী বড় নাকি $P(\{10\})$?

SOLUTION: এখানে $|S| = 10$. যদি $A = \{1, 4, 5\}$ নিই, তবে $|A| = 3$. সুতরাং $P(A) = \frac{|A|}{|S|} = \frac{3}{10}$. একই যুক্তিতে $P(\{1\}) = \frac{1}{10}$, এবং $P(\{10\})$-ও তাই। বস্তুতঃ যে কোনো singleton set-এরই probability সমান। সেই জন্যই তো "equally likely" কথাটা ব্যবহার করা হয়। ■

Example 19: Two boxes contain ten chips, each numbered 1 to 10. If one chip is drawn from each box, what is the probability that the sum of the numbers will be greater than 3?[5] **(2013.1b)**

SOLUTION: Probability-র অংকে মাঝেমাঝে একটা সমস্যা হয়--প্রশ্নকর্তা গুরুত্বপূর্ণ শর্তগুলো বলতে ভুলে যান, অথচ প্রত্যাশা করেন যে, ছাত্রছাত্রীরা শর্তগুলো আপনা থেকেই বুঝে ফেলবে। এটা খুবই দুঃখজনক। Probability যেহেতু অংকেরই একটা শাখা, সুতরাং প্রশ্নের ভাষায় কোনো ফাঁক থাকা উচিত নয়। যাই হোক, বলে যখন দেয় নি, অতএব শর্তগুলো কল্পনা না করে পথ নেই।

এখানে বলেছে যে দুটো বাক্স আছে, তাদের প্রত্যেকটায় দশটা করে কাগজের টুকরো আছে, 1 থেকে 10 সংখ্যা লেখা। প্রতিটা বাক্স থেকে একটা করে টুকরো নেওয়া হল। তারপর কিছু probability জানতে চাওয়া হয়েছে। কিন্তু এখানে তো কোনো random experiment-এর উল্লেখই নেই! আসলে বলা উচিত ছিল--"drawn randomly such that all outcomes are equally likely." এখানে আরও ধরে নিতে হবে যে, বাক্স দুটো distinct, মানে যেন "প্রথম বাক্স" আর "দ্বিতীয় বাক্স" বলে লেবেল করা আছে।

> Here the sample space S consists of all possible pairs of chips, one from each box,
>
> $$S = \{(i, j) \ : \ i, j \in \{1, ..., 10\}\}.$$

যেমন ধরো প্রথম বাক্স থেকে যদি 3 তোলা হয়, আর দ্বিতীয়টা থেকে 5, তবে তাকে লিখব $(3, 5)$.

> Thus $|S| = 10^2 = 100$.
>
> We assume that all these outcomes are equally likely.

এই শর্তটা বলা ছিল না, আমরা ধরে নিলাম। এবার দেখি আমাদের কোন event-টা নিয়ে কাজ করতে হবে--

> Let A denote the event
>
> $$A = \{(i, j) \in S \ : \ i + j > 3\}.$$

আমাদের এবার কাজ হল $|A|$ বার করা। যেহেতু 3 বেশ একটা ছোটো সংখ্যা, তাই অধিকাংশ (i,j)-র জন্যই $i+j>3$ হবে, ফলে A-টা বেশ বড় set. আমরা যদি A^c নিয়ে কাজ করি সেটা বরং অনেক ছোটো। তাই আমরা সেটার সাইজ বার করব, এবং সেখানে থেকে ঘুরপথে A-র সাইজ বার করব এইভাবে, $|A| = |S| - |A^c|$.

> Then
> $$A^c = \{(i,j) \in S \ : \ i+j \leq 3\} = \{(1,1),\ (1,2),\ (2,1)\}.$$
> Thus $|A| = |S| - |A^c| = 100 - 3 = 97.$
>
> Hence $P(A) = \frac{|A|}{|S|} = 0.97.$

■

Exercise 6: Two dice are thrown. Find the probability that the sum of the faces equals or exceeds 10.[3] **(2013.1c)**
HINT:
আগের অংকের অনুকরণে কর দেখি! ■

এর পরের অংকটার জন্য even (জোড়) আর odd (বিজোড়) সংখ্যাদের বিষয়ে কয়েকটা সহজ তথ্য লাগবে--

- দুটো even-এর গুণফল even-ই হয়, দুটো odd-এর গুণফল odd-ই হয়।
- যদি m,n দুটো integer হয় যাতে $m-n$ একটা even সংখ্যা, তবে হয় m,n দুজনেই even, নয়তো দুজনেই odd.
- যদি n একটা integer হয় তবে n এবং $|n|$ দুজনেই একইরকম হবে, মানে হয় দুজনেই even, নয়তো দুজনেই odd.

Even বা odd হওয়ার ব্যাপারটাকে অংকের ভাষায় বলে parity (যেমন ছেলে বা মেয়ে হওয়ার ব্যাপারটাকে বলে gender সেইরকম)। তাই দ্বিতীয় শর্তটাকে লিখতে পারি এইভাবে, যদি $m-n$ একটা even সংখ্যা হয়, তবে m,n-এর parity একই হতে বাধ্য। তৃতীয় শর্তটাকে লেখা যায় এইভাবে, n আর $|n|$-এর parity সর্বদা একই হয়।

Example 20: Two integers x and y are chosen at random with replacement from the first nine natural numbers $\{1,...,9\}$. Find the probability that $|x^2-y^2|$ is divisible by 2.[5] **(2005.1c)**
SOLUTION: এখানে একটা কথা ব্যবহার করা হয়েছে--"with replacement". এর মানেটা বুঝে নেওয়া যাক। মনে করো যেন তোমার সামনে একটা বাক্সে নয় টুকরো কাগজে $1,...,9$ লেখা আছে। তুমি একটা সংখ্যা চোখ বুঁজে টেনে নিলে। ধরো 2 পেলে। এবার যদি আরেকটা সংখ্যা টেনে নাও, তবে বুঝতেই পারছ যে, 2 আর আসবে না, কারণ বাক্সে খালি একটাই 2 ছিল, এবং সেটা তুমি প্রথম বারেই বাক্স থেকে তুলে নিয়েছ। এইভাবে তোলাকে বলে "without replacement". কিন্তু যদি তুমি প্রথম বার 2 পাওয়ার পর সেই 2-টাকে আবার বাক্সে ফিরিয়ে দিতে (মানে replace করতে) তবে দ্বিতীয়বারেও ফের 2 আসতে পারে। এইভাবে প্রত্যেকবার টানার পর সংখ্যাটাকে আবার বাক্সে ফিরিয়ে দেওয়া হলে তাকে বলে "with replacement".

> Let the sample space of the random experiment be S.
>
> Any two (not necessarily distinct) numbers x,y can be chosen in $9 \times 9 = 81$ ways.
>
> So $|S| = 81.$

> We assume that all these outcomes are equally likely.

এইবার event-টার দিকে তাকানো যাক।

> Let A be the event that $|x^2 - y^2|$ is even.

আমরা বার করতে চাই $P(A) = \frac{|A|}{|S|}$. এর মধ্যে $|S|$ তো বেরিয়েই গেছে। সুতরাং পড়ে আছে খালি $|A|$ বার করা। তার জন্য A-কে দুভাগে ভেঙে নিলে সুবিধা হবে।
বলেছে যে, $|x^2 - y^2|$-কে একটা even (জোড়) সংখ্যা হতে হবে। তার মানে $x^2 - y^2$ একটা জোড় সংখ্যা, আর তার মানে x^2, y^2 হয় দুজনেই even, নয়তো দুজনেই odd. আর এটা নিশ্চয়ই জানি যে, even সংখ্যার square সর্বদা even-ই হয়। একইভাবে odd-দের square হয় odd. সুতরাং হয় x, y দুজনেই even বা দুজনেই odd.

> $|x^2 - y^2|$ is divisible by 2
> $\iff x^2, y^2$ are both odd or both even
> $\iff x, y$ are both odd or both even.

এইবার একটু set দিয়ে লেখা যাক--

> Let A_1 be the event that both x, y are even. Let A_2 be the event that both x, y are odd.
>
> Then $A = A_1 \cup A_2$.
>
> $\because$ A number cannot be both odd and even at the same time,
>
> $\therefore A_1 \cap A_2 = \phi$.

সুতরাং $|A|$ বার করার জন্য $|A_1|$ আর $|A_2|$ আলাদা করে বার করে যোগ করে দিলেই হবে। সুতরাং পরবর্তী লক্ষ্য হল A_1, A_2-র সাইজ বের করা--

> Now A_1 can happen as follows:
>
> - Can choose even x in 4 ways ($\because$ exactly 4 even numbers in S).
> - Can choose even y again in 4 ways ($\because$ with replacement).
>
> So $|A_1| = 4 \times 4 = 16$.
>
> Similarly, A_2 can happen as follows:
>
> - Can choose odd x in 5 ways ($\because$ exactly 5 odd numbers in S).
> - Can choose odd y again in 5 ways ($\because$ with replacement).
>
> So there are $|A_2| = 5 \times 5 = 25$.
>
> So $|A| = |A_1| + |A_2| = 16 + 25 = 41$.
>
> So the required probability is $P(A) = \frac{|A|}{|S|} = \frac{41}{81}$.

■

Example 21: An integer is chosen at random from the set $\{1, 2, \ldots, 200\}$. What is the probability that the integer is divisible by 2 or 3?
(2007.1b)

SOLUTION:

> The sample space is
> $$S = \{1, 2, ..., 200\}.$$
> So $|S| = 200$.
> Let A_k be the event that the selected number is divisible by k.
> Then the required probability is $P(A_2 \cup A_3) = P(A_2) + P(A_3) - P(A_2 \cap A_3)$, by the inclusion-exclusion principle.
> Now $A_2 \cap A_3 = A_6$.

কারণ কোনো সংখ্যাকে 2 আর 3 দুটো দিয়েই ভাগ যায় একমাত্র তখনই যখন সংখ্যাটা 6 দিয়ে ভাগ যায়।

> Also for any positive integer k we have $|A_k| = [200/k]$. Here $[x]$ means the largest integer $\leq x$.

এইটা কেন হচ্ছে একটা উদাহরণ দিয়ে বুঝে রাখো-- ধরো $|A_3|$ বার করব, মানে 1 থেকে 200 পর্যন্ত কতগুলো সংখ্যা আছে যারা 3 দিয়ে ভাগ যায়। এরকম প্রথম সংখ্যা হল 3×1. সবচেয়ে বড় সংখ্যাটা বার করার জন্য $200/3$ বার করে দ্যাখো কত হয়। উত্তর হল $66\frac{2}{3}$. তার মানে A_3-এর মধ্যে সবচেয়ে বড় সংখ্যাটা হল 3×66. দেখলে আমরা কীভাবে ভগ্নাংশটুকু বাদ দিলাম, মানে $[200/3]$ ব্যবহার করলাম। দেখাই যাচ্ছে যে $A_3 = \{3 \times 1, 3 \times 2, ..., 3 \times 66\}$, তাই এর সাইজ 66 হবে, মানে $|A_3| = [200/3]$.

> So
> $$\begin{aligned} P(A_2) &= \frac{|A_2|}{200} = \frac{100}{200}, \\ P(A_3) &= \frac{|A_3|}{200} = \frac{66}{200}, \\ P(A_6) &= \frac{|A_6|}{200} = \frac{33}{200}. \end{aligned}$$

এখানে ভগ্নাংশগুলোকে কাটাকাটি না করাই ভালো। সবারই নীচে একই সংখ্যা 200 থাকায় পরে যোগ-বিয়োগ করতে সুবিধা হবে।

> So the required probability is
> $$P(A_2) + P(A_3) - P(A_6) = \frac{100}{200} + \frac{66}{200} - \frac{33}{200} = \frac{133}{200}.$$

■

Example 22: What is the probability that a leap year will contain 53 Sundays? **(2006.1e)**

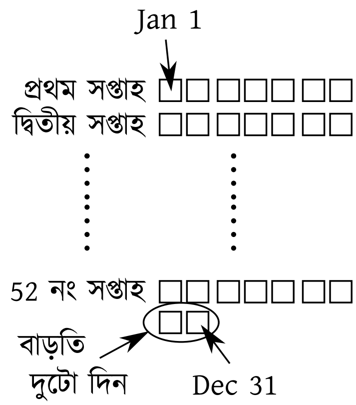


Fig 16

SOLUTION: এটা একটা অত্যন্ত জঘন্যভাবে লেখা অংক। এখানে মাথামুণ্ডু কিছুই বোঝা যাচ্ছে না যে probability-র প্রশ্নটা আসছে কোথা থেকে। আসলে এখানে কল্পনা করা হচ্ছে যেন random-ভাবে একটা leap year এমনভাবে নির্বাচন করা হচ্ছে, এবং তার পয়লা জানুয়ারী কোন বারে পড়েছে সেটা দেখা হচ্ছে। এইটা হল আমাদের random experiment. আরও ধরে নেওয়া হচ্ছে যেন এর প্রতিটা outcome-ই (মানে সাতটা বারের প্রত্যেকটা) equally likely. এই কথাটা আমরা গোড়াতেই লিখে নেব।

> We assume that the first day of a randomly selected leap year is equally likely to be any day of the week.
> Then the sample space is $S = \{$Sun, Mon, Tue, Wed, Thu, Fri, Sat$\}$.
> So $|S| = 7$.

এইবার event-টার হিসেব নেওয়া যাক--

> A leap year consists of $366 = 52\times 7 + 2$ days. Thus,all the days occur exactly 52 times, except the days of the week for Jan 1 and Jan 2, which occur 53 times each.

স্পষ্ট না হলে Fig 16 দেখে নাও। যেকোনো leap year-এই 52-টা আস্ত সপ্তাহ থাকেই, সুতরাং 52-টা রবিবার অন্ততঃ থাকবেই। যদি 53-টা রবিবার থাকতে হয়, তবে সেই 53 নম্বর রবিবারটাকে শেষের বাড়তি দিন দুটোর মধ্যে একটা হতে হবে। সেই দিন দুটোর বার আর বছরের প্রথম দিনদুটো বার একই (ছবিতে ওরা একই vertical লাইনে আছে)।

> Thus there can be 53 Sundays if and only if Jan 1 or 2 is a Sunday, *i.e.*, if and only if the leap year starts on a Sunday or a Saturday.
> So our event is $A = \{$ Sun, Sat $\}$.
> So the required probability is $P(A) = \frac{|A|}{|S|} = \frac{2}{7}$.

■

DAY 4
Equally likely outcomes (part 2)

বুঝতেই পারছ যে এই অংকগুলোতে খালি দুটো set-এর সাইজ বার করতে পারলেই হয়, sample space-এর এবং যে event-এর probability চাইছ সেটার। কোনো set-এর সাইজ বার করার অন্যতম হাতিয়ার হল permutation আর combination লাগানো। তাই এই সুযোগে permutation-combination-এর অংকগুলোকে probability-র অংক বলে চালিয়ে দেওয়া যায়। Permutation-combination-এর প্যাঁচের উপর নির্ভর করে এই অংকগুলো ভয়ানক কঠিন হতে পারে, এবার আমরা এরকম কয়েকটা অংক দেখব। এগুলোতে probability-র জ্ঞান অতি সামান্যই লাগবে। এই নিতান্ত পরীক্ষা-মার্কা অংকগুলো কঠিন লাগা স্বাভাবিক, তাতে ভয় পেও না। বাস্তবে এইসব অংকের প্রয়োগ নেই বললেই হয়, এই বইতেও এগুলো পরে কোথাও লাগবে না। কাল থেকেই আমরা আবার সহজ জিনিসে ফিরে যাবো।

4.1 Finite sample space-এর আরও অংক (একটু কঠিন)

এখানে যেটুকু permutation আর combination লাগবে, সেটা হায়ার সেকণ্ডারী থেকেই জানা থাকার কথা। চাইলে আমাদের বাংলায়-বোঝানো-ইংরাজি-বই সিরিজের Permutation and Combination বইটা দেখতে পারো।

Example 23: From the numbers $1, 2, ..., (2n + 1)$, three numbers are chosen at random. Prove that the probability that these are in arithmetical progression is $\frac{3n}{(4n^2-1)}$.[5] **(2012,2009,2004)**

SOLUTION: এরকম অংক প্রথমে দেখেই একটু হকচকিয়ে যেতে হয়। কী করে এগোতে হবে ফস্ করে বোঝা যায় না। এসব ক্ষেত্রে চিন্তা করার একটা ভালো কায়দা হল সহজ কিছু উদাহরণ নিয়ে দেখা। মনে কর $n = 3$ নিলাম। তাহলে $2n + 1 = 7$, অর্থাৎ আমাদের কাছে সাতটা সংখ্যা আছে--

$$1, 2, 3, 4, 5, 6, 7.$$

এদের থেকে যেকোনো তিনটে নিতে হবে। কতভাবে নেওয়া যায়? হায়ার সেকণ্ডারীর স্মৃতি থেকে তোমার মাথায় চট্ করে উত্তর আসবে $^7C_3$. আর তার কিছুক্ষণের মধ্যেই মনের মধ্যে একটা খোঁচা অনুভব করবে--নাকি $^7P_3$? হায়ার সেকণ্ডারীর করুণ স্মৃতি তোমার কানে ফিস্‌ফিস্ করে বলবে যে, এ সমস্যা কোনোদিনই সমাধান করা যাবে না। কারণ অংকে তো কোথাও বলে দেয় নি যে, এখানে order-টা গুরুত্বপূর্ণ কিনা, অর্থাৎ $1, 3, 6$ নেওয়া আর $3, 6, 1$ নেওয়া কী একই ব্যাপার নাকি আলাদা! ওই দুশ্চিন্তা আপাততঃ ভুলে যাও। উদাহরণটা নিয়ে কাজ করতে থাকো--ধরো তিনটে সংখ্যা তুলেছো, $5, 1, 3$. এরা কি AP-তে আছে? এখানেও সেই order-এর প্রশ্নটা ফিরে আসছে। যদি প্রথমে 5, তারপরে 1 আর সবশেষে 3--এইভাবে ভাবো, তবে AP-তে নেই। কিন্তু যদি নিজের ইচ্ছে মত সাজিয়ে নিতে পারো, তবে $1, 3, 5$ ভাবলে AP-তে আছে। দুঃখের কথা "ইচ্ছে মত সাজিয়ে নেওয়া"-টা চলবে কিনা সেটা প্রশ্নে কিছু বলে নি। কিন্তু তার জন্য তোমার কোনো দোষ নেই। দোষটা সম্পূর্ণই প্রশ্নের, যাবতীয় শর্ত স্পষ্ট করে বলে নি। এমতাবস্থায় উচিত হল নিজের সুবিধা মত কোনো একটা ব্যাখ্যা ধরে নিয়ে দেখা উত্তরটা মেলে কিনা। না মিললে ফের অন্য ব্যাখ্যা চেষ্টা করা যাবে। তা আমরা এই ব্যাখ্যা ধরে এগোই--order গুরুত্বপূর্ণ নয়। অতএব ইচ্ছে মত সাজিয়ে নিলে আপত্তি নেই। তাহলে $|S| = {}^7C_3$. আর AP হবে খালি এই কয়টা ক্ষেত্রে

$$1, 2, 3; \quad 2, 3, 4; \quad 3, 4, 5; \quad 4, 5, 6; \quad 5, 6, 7;$$

এদের ক্ষেত্রে common difference হল 1

$$1, 3, 5; \quad 2, 4, 6; \quad 3, 5, 7;$$

এদের ক্ষেত্রে common difference হল 2

$$1, 4, 7;$$

এক্ষেত্রে common difference হল 3

তার মানে মোট 9 ক্ষেত্রে। লক্ষ কর যে common difference এখানে 3-এর বেশী হতে পারে না, কারণ তাহলে 1 থেকে 7-এর মধ্যে তিনটে সংখ্যা ধরবে না। তাহলে probability দাঁড়াচ্ছে $9/{}^7C_3 = \frac{9}{35}$. যেটা প্রমাণ করতে দিয়েছে মানে

$\frac{3n}{(4n^2-1)}$, সেখানে $n=3$ বসিয়ে দ্যাখো, একই উত্তর পাবে। যাক, আমাদের ব্যাখ্যাটা অন্ততঃ $n=3$-এর জন্য মিলেছে। এবার দেখি order যদি গুরুত্বপূর্ণ হত তবে কী উত্তর হত। সেক্ষেত্রে নীচের তলায় থাকত ${}^7P_3$ কিন্তু উপরের তলাতে এবার হবে $2\times 9=18$, কারণ $1,3,5$-এর সঙ্গে এবার $5,3,1$-কেও আলাদা করে গুণতে হবে। এভাবে প্রত্যেকটার জন্যই একটা করে উল্টো সংস্করণ পাবে। সেই জন্যই 2 দিয়ে গুণ করতে হল। তাই উত্তর হত $18/{}^7P_3=\frac{3}{35}$, যেটা মিলছে না। সুতরাং মনে হচ্ছে প্রথম ব্যাখ্যাটা ধরে এগোনোই ভালো।

এইখানে আরেকবার মনে করিয়ে দিই যে, এই যে উদাহরণ নিয়ে আন্দাজ করতে হল, সেটা কিন্তু probability শেখার অংশ নয়। সেটা নিতান্তই প্রশ্নের অসম্পূর্ণতার দোষে। দুঃখের ব্যাপার, প্রশ্নের এইরকম গাফিলতির ফলে probability বিষয়টা যতটা না কঠিন, তার চেয়ে অনেক বেশী কঠিন বলে বোধ হয়।

যাই হোক, এবার তবে সমাধানটা লিখতে শুরু করি।

> Here the sample space, S, consists of all possible ways of choosing 3 numbers from the $2n+1$ numbers, assuming that order is not important.
>
> So $|S|=\binom{2n+1}{3}$.

এই notation-টা নিশ্চয়ই জানো যে, $\binom{n}{r}$ মানে ${}^nC_r$.

> Let A be the event that the chosen numbers are in an AP.
>
> The common difference can be any integer from 1 to n.

কারণ তার থেকে বেশী হলে 1 থেকে n-এর মধ্যে তিনটে সংখ্যা ধরবে না।

> Thus
> $$A=\cup_{k=1}^n A_k,$$
> where A_k is the event that the chosen numbers form an AP with common difference k.
>
> Clearly, A_k's are pairwise disjoint.

কারণ একই AP-র দুটো আলাদা common difference হয় না।

> $\therefore |A|=\sum_1^n |A_k|$.

এবার আমরা $|A_k|$-গুলো বার করব। A_k-র element-রা দেখতে কেমন একটু ভেবে নিই। যেমন $n=3$ নিয়ে যে উদাহরণটা দেখেছিলাম সেখানে A_2-র element-রা ছিল $(1,3,5)$, $(2,4,6)$ আর $(3,5,7)$.

> Any element of A_k is a triple of the form $(x,x+k,x+2k)$. Since all these are from $\{1,...,2n+1\}$, so $x\in\{1,...,2n+1-2k\}$.

কারণ $x+2k\leq 2n+1$ চাই।

> Thus $|A_k|=2n+1-2k$.

এবার $|A_k|$-গুলোকে যোগ করে $|A|$ বার করব--

Hence

$$\begin{aligned} |A| &= \sum_1^n |A_k| \\ &= \sum_1^n (2n+1-2k) \\ &= n(2n+1) - 2\sum_1^n k \\ &= n(2n+1) - n(n+1) \\ &\qquad \text{কারণ } 1+2+\cdots+n = \tfrac{n(n+1)}{2} \\ &= n^2. \end{aligned}$$

So

$$P(A) = \frac{|A|}{|S|} = \frac{n^2}{\binom{2n+1}{3}} = \frac{6n^2}{(2n+1)(2n)(2n-1)} = \frac{3n}{(4n^2-1)},$$

as required.

■

Example 24: If a die is thrown k times show that the probability of an even number of sixes is

$$\left\{1 + \left(\tfrac{2}{3}\right)^k\right\}/2.$$

[5] **(2011,2008,2006)**

SOLUTION: প্রথমে একটা উদাহরণ দিয়ে বুঝে নিই যে, কী চেয়েছে। ধরো একটা ছক্কা আছে, সেটাকে 3 বার চালা হয়েছে। তবে কতবার ছয় পড়তে পারে? উত্তর হল--0 বা 1 বা 2 বা 3-বার। আমরা জানতে চাই even (জোড়) সংখ্যকবার ছয় পড়ার probability. তার মানে আমাদের উদাহরণে 0 বা 2-বার[3]।
এখানে আমরা ধরে নেব যে, ছক্কাটার সব দিক সমান ভারী, আর throw-গুলো পরস্পরের সঙ্গে independent.

We assume that the die is fair, and that the throws are mutually independent.

এখানে "mutually independent" কথাটা একটা technical term, যার বিশদ সংজ্ঞা কালকে দেব। আপাততঃ এর মানেটা সহজ ভাষায় জেনে রাখো--ছক্কার একটা চালে কী পড়ল তার সাথে অন্য চালগুলোর কোনো সম্পর্ক নেই (যেটাই বলাই বাহুল্য!)।

The sample space S is

$$S = \{(x_1, ..., x_k) \ : \ \forall i \quad x_i \in \{1, ..., 6\}\}.$$

[3] হ্যাঁ, 0 একটা even সংখ্যা। সে নিয়ে কোনো বিতর্ক নেই। যদিও অনেক ছাত্রছাত্রীর মনেই এ নিয়ে একটা ধন্দ থাকে দেখেছি।

যেমন আমাদের $k=3$ বার ছক্কা চালার উদাহরণে যদি প্রথমে ছয়, তারপরে পুট এবং সবশেষে আবার পুট পড়ে, তবে সেই outcome-টাকে লিখব $(6,1,1)$. এইভাবে প্রতিটা outcome-ই দেখতে হবে (x_1,x_2,x_3)-এর মত, যেখানে x_1,x_2,x_3 প্রত্যেকেই 1 থেকে 6-এর মধ্যে কিছু একটা সংখ্যা।

> Clearly $|S|=6^k$.

কারণ x_1-এর 6 রকম value হতে পারে, x_2-এরও তাই, x_3-এরও তাই।

> Let A be the event that an even number of sixes has occured.

এবার $|A|$ বার করতে হবে। তার জন্য A-কে কয়েকটা ভাগে ভেঙে নেব, কতবার 6 পড়েছে তার ভিত্তিতে--

> Then
> $$A=\bigcup_{r \text{ even}} A_r,$$
> where A_r is the event that exactly r sixes have occured.
> Clearly, A_r's are pairwise disjoint. So $|A|=\sum_{r \text{ even}}|A_r|$.

এইবার আমরা A_r-গুলোর সাইজ বার করব। সেই $k=3$-বার ছক্কা চালার উদাহরণ দিয়ে চিন্তা করে নিই। সেখানে A_1-এর সাইজ কত? A_1 মানে ঠিক একবারই ছয় পড়েছে। সেই ছয়টা যদি প্রথম দানেই এসে থাকে তবে outcome-টা এই জাতীয় হবে-- $(6,\neq 6,\neq 6)$. যদি দ্বিতীয়দানে এসে থাকে তবে হবে $(\neq 6, 6, \neq 6)$-এর মত, ইত্যাদি। এইভাবে "কোন দানে ছয় পড়েছে" তার ভিত্তিতে তিনরকম কেস পাচ্ছি। প্রত্যেকটা "$\neq 6$"-র জায়গায় 5 রকম সংখ্যা সম্ভব। দুটো করে "$\neq 6$" আছে, তাই মোট 5×5 ভাবে হতে পারে। এরকম তিনটে কেস, তাই উত্তর হবে 3×5^2. এই উদাহরণটা ভালো করে বুঝে নিয়ে নীচের জিনিসটা পড়ো।

> Then $|A_r|$ can be computed as follows.
>
> 1. Select r out of k throws for the sixes: $\binom{k}{r}$ ways.
> 2. Take numbers from $\{1,...,5\}$ for the remaining $k-r$ throws: 5^{k-r} ways.
>
> So $|A_r|=\binom{k}{r}5^{k-r}$.

এবার A_r-দের জুড়ে জুড়ে A বানাতে হবে--

> $$\begin{aligned} |A| &= \sum_{r \text{ even}} |A_r| \\ &= \sum_{r \text{ even}} \binom{k}{r} 5^{k-r}. \end{aligned}$$

এই চেহারাটা দেখেই তোমার binomial theorem-এর কথা মনে পড়া উচিত। নীচের ধাপ কটা আমরা হায়ার সেকণ্ডারীতেই শিখি (বা অন্ততঃ শেখা উচিত)। মনে না থাকলে এখান থেকে নতুন করে শিখে নাও। নয়তো আমাদের বাংলায়-বোঝানো-ইংরাজি-বই সিরিজের Permutation and Combination বইটা দেখতে পারো।

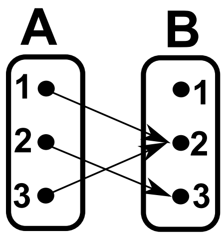


Fig 17

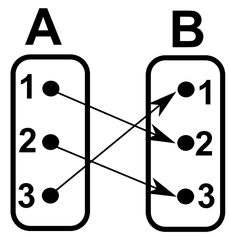


Fig 18

Now, from the binomial theorem,

$$(5+1)^k = \sum_{r=0}^{k} \binom{k}{r} 1^r 5^{k-r},$$

$$(5-1)^k = \sum_{r=0}^{k} \binom{k}{r} (-1)^r 5^{k-r}.$$

Adding

$$(5+1)^k + (5-1)^k = 2 \sum_{r \text{ even}} \binom{k}{r} 5^{k-r},$$

or

$$\sum_{r \text{ even}} \binom{k}{r} 5^{k-r} = \frac{1}{2}(6^k + 4^k).$$

So

$$P(A) = \frac{|A|}{|S|} = \frac{1}{2} \times \frac{6^k + 4^k}{6^k} = \left\{1 + \left(\tfrac{2}{3}\right)^k\right\}/2,$$

as required.

■

Example 25: One mapping is selected at random from all the mappings of the set $A = \{1, 2, ..., n\}$ into itself. Find the probability that the mapping selected is one-one.[5] **(2010.2e)**

SOLUTION: যদি one-one মানে কী ভুলে গিয়ে থাকো, তবে Fig 17 আর Fig 18 দেখে নাও। Fig 17-এর function-টা one-one নয়, কারণ দুটো তীর একই জায়গায় (2-তে) পৌঁছেছে। কিন্তু Fig 18-এর function-টা one-one. এই অংকের ভাষায় একটু সমস্যা আছে ওই "into" শব্দটা নিয়ে। সাধারণতঃ একটা mapping-কে "into" বলা হলে বুঝতে হবে যে সেটা "onto" নয়। কিন্তু এখানে সেই অর্থে ব্যবহার করা হয়নি। তাই "onto" বা "into" কোনোকিছুই না বলে খালি "to" বললেই ভালো করত।

Here the sample space S consists of all mappings from $\{1, ..., n\}$ to itself. Since for any such mapping f, each $f(i)$ can have n possible values, so $|S| = n^n$.
Let A be the event that the randomly selected mapping is one-one.

If $f : \{1, ..., n\} \to \{1, ..., n\}$ is one-one, then

1. $f(1)$ can take n possible values,
2. $f(2)$ can take $n-1$ possible values, since $f(2) \neq f(1)$,
3. $f(3)$ can take $n-2$ possible values, since $f(3) \neq f(1), f(2)$,

$$\vdots$$

n. Finally, $f(n)$ can take $n-(n-1) = 1$ possible value, since $f(n) \neq f(1), f(2), ..., f(n-1)$.

So $|A| = n \times (n-1) \times \cdots \times 1 = n!$.
Hence the required probability is

$$P(A) = \frac{|A|}{|S|} = \frac{n!}{n^n} = \frac{(n-1)!}{n^{n-1}}.$$

■

Example 26: From a pack of 52 cards an even number of cards are drawn. Find the probability that these consist half of red and half of black. **(2006.1c)**

SOLUTION: এই অংকটা তাস নিয়ে। সুতরাং এক প্যাকেট তাসে কী কী থাকে একটু জানা থাকা দরকার। একটা প্যাকেটে 52-টা তাস থাকে। তাদের মধ্যে 26-টা লাল তাস, আর বাকি 26-টা কালো। ব্যস্ এটুকু জানলেই এই অংকটা করা যাবে। তবে আরও জেনে রাখা ভালো যে, লাল তাসগুলোর মধ্যে 13-টা হল heart (হরতন) আর বাকি 13-টা diamond (রুইতন)।

Fig 19

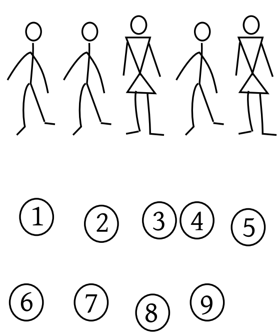
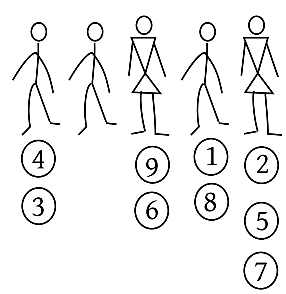


Fig 20



কালোদের বেলায় 13-টা spade (ইস্কাবন) আর বাকি 13-টা club (চিঁড়েতন)।

> Let $2k$ cards be drawn. Then the sample space S consists of all the possible ways of drawing $2k$ cards from 52 cards.
>
> So $|S| = \binom{52}{2k}$.
>
> Let A be the event where k red and k black cards are selected.
>
> Then $|A| = \binom{26}{k} \times \binom{26}{k}$.

যেহেতু 26-টা লাল তাস থেকে k-টা নেওয়া হচ্ছে, আর 26-টা কালো থেকে বাকি k-টা।

> Thus the required probability is
>
> $$P(A) = \frac{|A|}{|S|} = \frac{\binom{26}{k}^2}{\binom{52}{2k}}.$$

■

Example 27: If n objects are distributed among a men and b women $(b < a)$, show that the probability that the women get an odd number of objects is

$$\frac{1}{2}[(a+b)^n - (a-b)^n]/(a+b)^n.$$

(2008.1d)

SOLUTION: একটা উদাহরণ নিলে বুঝতে সুবিধা হবে। Fig 19-এ $a = 3$ জন ছেলে আর $b = 2$ জন মেয়ে নিয়েছি। আর জিনিস রয়েছে $n = 9$ খানা। ধরো এদেরকে Fig 20-এর মত করে ভাগ করে দিয়েছি। তাতে প্রথম ছেলেটা 2-টো পেয়েছে, দ্বিতীয় ছেলেটার বরাতে কিছুই জোটেনি, এইরকম। মেয়েরা মোট পেয়েছে $2 + 3 = 5$-টা। প্রশ্ন হল মেয়েদের পাওয়া মোট জিনিসের সংখ্যা odd হওয়ার probability কত।

> Here we assume that the objects are distinct, each object may go to any person, and all the possible ways of doing this are equally likely.
>
> As there are n objects, and each object may go to any of the $a+b$ persons, so the total number of ways is $(a+b)^n$.
>
> Let A be the event that the women get an odd number of objects. Then
>
> $$A = \bigcup_{k \text{ odd}} A_k,$$
>
> where A_k is the event that exactly k objects go to the women. Clearly

> A_k's are pairwise disjoint, and so
>
> $$|A| = \sum_{k \text{ odd}} |A_k|.$$

আমরা এবার $|A_k|$-গুলো আলাদা করে বার করে যোগ করে দিলেই $|A|$ পেয়ে যাব।

> $|A_k|$ may be computed as follows.
>
> 1. Select the k objects for the women: $\binom{n}{k}$ ways.
>
>   এখানে Fig 20-এর মত $k = 5$ নিয়ে ভাবো। 9-টা থেকে কোন 5-টা জিনিস মেয়েরা পাবে সেটা ঠিক করা যাবে $\binom{9}{5}$ কায়দায়।
> 2. Distribute these k objects among the women: b^k ways.
>
>   এই 5-টা জিনিস কীভাবে মেয়েদের মধ্যে ভাগাভাগি হচ্ছে।
> 3. Distribute the remaining $n-k$ objects among the men: a^{n-k} ways.
>
> So $|A_k| = \binom{n}{k} a^{n-k} b^k$.

আবার binomial theorem-এর চেহারাটা এসে গেছে।

> Now using the binomial theorem
>
> $$\begin{aligned}(a+b)^n &= \sum_k \binom{n}{k} a^{n-k} b^k, \\ (a-b)^n &= \sum_k (-1)^k \binom{n}{k} a^{n-k} b^k.\end{aligned}$$
>
> So
>
> $$\frac{(a+b)^n - (a-b)^n}{2} = \sum_{k \text{ odd}} \binom{n}{k} a^{n-k} b^k.$$
>
> So
>
> $$P(A) = \frac{|A|}{|S|} = \frac{(a+b)^n - (a-b)^n}{2(a+b)^n},$$
>
> as required.

■

Exercise 7: If n biscuits are distributed among n beggars, what is the chance that a particular beggar receives $r(< n)$ biscuits?[6]
(2005.2c)
HINT:
আগের অংকটার অনুকরণে কর দেখি! ■

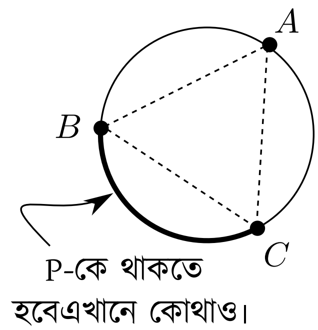


Fig 21

4.2 Infinite sample space

এতক্ষণ আমরা finite সাইজের sample space-এর উপর "equally likely outcomes" নিয়ে কাজ করছিলাম। অংকের ভাষায় এর অর্থ ছিল যদি S হয় sample space-টা, আর A হয় কোনো একটা event, তাহলে $P(A) = \frac{|A|}{|S|}$. এই যে আমরা $|S|$ দিয়ে দিব্যি ভাগ করে দিতে পারলাম, তার কারণ S ছিল finite এবং nonempty. যদি S একটা infinite set হয়, তবে এরকমভাবে ভাগ করা যেত না। কিন্তু তাহলেও কখনও কখনো একই ধরণের একটা জিনিস করা যায়। এরকমই একটা উদাহরণ আসছে এবার।

Example 28: A point P is chosen at random on a circle of radius a, and A is a fixed point on the circle. Find the probability that the chord AP will exceed the length of the side of an equilateral triangle inscribed in the circle.[5] **(2008.2a)**

SOLUTION: এখানে random experiment-টা হল একটা circle-এর উপর থেকে (মানে তার circumference বা পরিধি বরাবর) কোনো একটা point নেওয়া। সুতরাং sample space হল পুরো circle-টাই। মনে রেখো যে অংকের জগতে circle বলতে খালি তার circumference বা পরিধিটুকুই বোঝায়, ভিতরের অংশটা নয় (সেটা বোঝাতে চাইলে disc শব্দটা ব্যবহার করা হয়)। সুতরাং S এখানে গোটা circle-টার যাবতীয় point-এর set, যেটা একটা infinite set. এখানে আমরা এই set-এর থেকে random-ভাবে একটা point নিচ্ছি, কোনো দিকে কোনো পক্ষপাত না দেখিয়ে, প্রতিটা point-কেই সমান সুযোগ দিয়ে। অংকের ভাষায় এর অর্থ হল--যদি $A \subseteq S$ কোনো event হয় তবে $P(A) = \frac{\text{length}(A)}{\text{length}(S)}$ হবে। এখানে $\text{length}(S)$ মানে S-এর দৈর্ঘ্য, অর্থাৎ $2\pi a$. একইভাবে $\text{length}(A)$ মানে হল A-র দৈর্ঘ্য। লক্ষ কর আমরা কীভাবে $|A|$ আর $|S|$-এর বদলে $\text{length}(A)$ আর $\text{length}(S)$ নিয়ে কাজ করছি। গোড়াতেই এই কথাটা লিখে শুরু করব--

> Since P is chosen at random, so the probability that it lies in any given arc M is
> $$\frac{\text{length}(M)}{\text{length}(S)} = \frac{\text{length}(A)}{2\pi a}.$$

এইবার আসল ধাপগুলো। একটু ছবি এঁকে নিলে সুবিধা হবে (Fig 21)।

> A is a given point on the circle.
>
> Let B, C be points on the circle such that ABC is an equilateral triangle.
>
> Clearly, by symmetry, the arcs $\widehat{AB}, \widehat{BC}, \widehat{CA}$ are of equal lengths.

From the the geometry it is clear that the chord $\overline{AP}$ will exceed the length of a side of the triangle if and only if P lies on the arc $\widehat{BC}$.
Thus, the required probability is the probability that P lies on the arc $\widehat{BC}$.
Since the arc $\widehat{BC}$ is $\frac{1}{3}$ of the total circumference, hence the required probability is

$$\frac{\text{length}(\widehat{BC})}{2\pi a} = \frac{1}{3}.$$

■

DAY 5
Independence

Independence-এর ধারণাটা আমরা এই অধ্যায়ের গোড়ার দিকেই উল্লেখ করেছিলাম। যদি পরপর দুবার কয়েন টস কর তবে তারা independent হবে, অর্থাৎ কিনা একটা টস অন্যটাকে প্রভাবিত করতে পারে না। এটা বলাই বাহুল্য। এবার আমরা এই ধারণাটাকে অংকের ভাষায় প্রকাশ করব।
আমরা এখানে দুটো কয়েন টসের মধ্যে independence নিয়ে আলোচনা করব, তাই সুবিধার জন্য random experiment-টা নেব "পরপর দুটো কয়েন টস করা"। এই আলোচনার জন্য দুবার দুটো আলাদা কয়েন নিলেও আমাদের আপত্তি নেই। ধরো প্রথম কয়েনটার $P(\text{head}) = 0.6$ আর দ্বিতীয়টার $P(\text{head}) = 0.5$. তার মানে যদি A, B দুটো event নিই, যেখানে A মানে হল প্রথম কয়েনে head পড়া (দ্বিতীয় কয়েনে যা খুশি), এবং B হয় দ্বিতীয় কয়েনে head পড়া (প্রথম কয়েনে যা খুশি)। অংকের ভাষায়, sample space-টা হল

$$S = \{(H, H), \quad (H, T), \quad (T, H), \quad (T, T)\},$$

আর

$$\begin{aligned} A &= \{(H, H), \quad (H, T)\}, \\ B &= \{(H, H), \quad (T, H)\}. \end{aligned}$$

তার মানে random experiment-টা বহুবার করলে মোটামুটি 60% বার A ঘটবে, আর 50% বার B ঘটবে। চট করে বলো তো দুটো কয়েনেই একই সঙ্গে head পড়ার probability কত হবে? প্রথম কয়েনে head তো 60% বার আসছিলই, তার মধ্যে নিশ্চয়ই 50% ক্ষেত্রে দ্বিতীয় কয়েনেও head আসবে। সুতরাং দুটোতেই head পড়ার probability দাঁড়াচ্ছে 60%-এর $50\% = 30\%$.
এই যে probability দুটো গুণ হয়ে গেল, এখান থেকেই independence-এর সংজ্ঞাটা আসে--

Example 29: Define independence of two events.[1] **(2013,2009)**

SOLUTION:

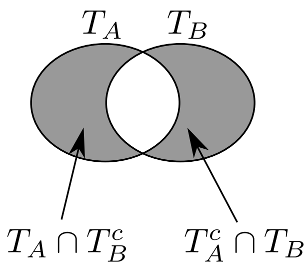


Fig 22

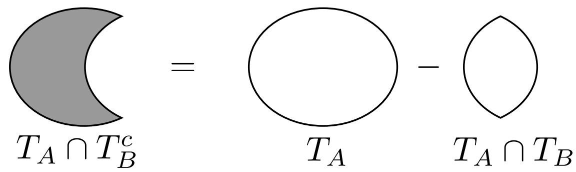


Fig 23

> **DEFINITION: Independence of two events**
>
> Let A, B be two events connected with the same random experiment. Then they are called independent if $P(A \cap B) = P(A)P(B)$.

■

Example 30: A speaks the truth in 75% cases, B speaks the truth in 80% cases. What is the probability that they will contradict each other?
(2008.1e)
SOLUTION:

We assume

- we are dealing with a question with yes/no answer. Thus if both the speakers lie, then the will not contradict each other.
- A and B behave independently,

Let C be the event that they contradict each other.

Let T_A be the event that A speaks the truth, and T_B be the event that B speaks the truth.

Given that $P(T_A) = 0.75$ and $P(T_B) = 0.8$.

They will contradict each other if and only if exactly one of them speaks the truth, ie, $C = (T_A \cap T_B^c) \cup (T_A^c \cap T_B)$.

এটা কী করে এল? ওরা পরস্পরবিরোধিতা করবে যদি

- A সত্যি বলে, কিন্তু B মিথ্যে বলে $(T_A \cap T_B^c)$ এইটা হল Fig 22-এর বাঁদিকের চাঁদটা,

- আর নয়তো A মিথ্যে বলে, কিন্তু B সত্যি বলে $(T_A^c \cap T_B)$, মানে Fig 22-এর ডানদিকের চাঁদটা।

The probability of this is

$$\begin{aligned} P(C) &= P(T_A \cap T_B^c) + P(T_A^c \cap T_B) \quad [\because \text{disjoint}] \\ &= \big[P(T_A) - P(T_A \cap T_B)\big] + \big[P(T_B) - P(T_A \cap T_B)\big] \\ &\qquad \text{কারণটা বুঝবে Fig 23 দেখলেই।} \\ &= 0.75 + 0.8 - 2 \times 0.75 \times 0.8 = 0.35. \end{aligned}$$

■

Example 31: If any two events A and B are independent, then show that A', B' as well as A, B' are also independent (where '′' means complement).[4]
(2014.1b)
SOLUTION: এখানে যেটা প্রমাণ করতে দিয়েছে সেটা খুবই সহজ বুদ্ধির কথা। ধরো একটা fair কয়েনকে পরপর দুবার টস করবে। প্রথম টসে head এল। এইটুকু শুনে কিছু আন্দাজ করতে পারো পরের টসে head আসার সম্ভাবনা কত? উত্তর হল $\frac{1}{2}$, এবং সে কথাটা প্রথম টসে head পড়ার আগেই জানতাম। তার মানে "প্রথম টসে head পড়েছে" আর "দ্বিতীয় টসে head পড়বে" এই event দুটো পরস্পরের সঙ্গে independent. আচ্ছা যদি প্রথম টসে tail পড়ত? তাহলেও একই গল্প খাটত, কারণ head পড়া আর tail পড়া তো একই জিনিসের দুটো পিঠ। একই যুক্তিতে বোঝা যায় যে, A, B যদি independent হয়, তবে A^c, B-ও তাই হবে। A, B^c এবং A^c, B^c-এর বেলাতেও একই কথা বলা যাবে।

$$\begin{aligned} P(A^c \cap B^c) &= P\big((A \cup B)^c\big) \\ &= 1 - P(A \cup B) \\ &= 1 - P(A) - P(B) + P(A \cap B) \\ &= 1 - P(A) - P(B) + P(A)P(B) \quad [\because A, B \text{ are independent}] \\ &= (1 - P(A))(1 - P(B)) \\ &= P(A^c)P(B^c). \end{aligned}$$

So A^c, B^c are independent.
We have $A = (A \cap B) \cup (A \cap B^c)$.
Also $A \cap B$ and $A \cap B^c$ are disjoint.
So $P(A) = P(A \cap B) + P(A \cap B^c)$.
Hence $P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B)$, since A, B are independent.
Thus

$$P(A \cap B^c) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c).$$

So A, B^c are independent.

■

Exercise 8: Show that if two events A and B are independent then $\bar{A}$ and $\bar{B}$ are also independent.[2]
(2007,2005)
HINT:
আগের অংকেই আছে। ■

Example 32: Let $A_1, A_2, ..., A_n$ be n events connected to a random experiment. Define mutual independence and pairwise independence. Give a suitable example to show that pairwise independence does not always imply mutual independence.[1+1+2] **(2013,2009)**
SOLUTION:

DEFINITION: **Mutual and pairwise independence**

Let a random experiment have sample space S. Let $A_1, ..., A_n \subseteq S$ be some events. They are called

- mutually independent if for every subset $\{i_1, ..., i_r\} \subseteq \{1, ..., n\}$

$$P(A_{i_1} \cap \cdots \cap A_{i_r}) = P(A_{i_1}) \times \cdots \times P(A_{i_r}).$$

- pairwise independent if

$$\forall i \neq j \in \{1, ..., n\} \quad P(A_i \cap A_j) = P(A_i) \times P(A_j).$$

Pairwise independence does not imply mutual independence:

Let our random experiment be tossing a fair coin three times independently. Let the outcomes be (X, Y, Z).

$$\begin{aligned} A_1 &= \text{the event that } X = Y \\ A_2 &= \text{the event that } Y = Z \\ A_3 &= \text{the event that } Z = X. \end{aligned}$$

Then $P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}$.
Also $P(A_1 \cap A_2) = P(A_2 \cap A_3) = P(A_3 \cap A_1) = \frac{1}{4}$.
Also $P(A_1 \cap A_2 \cap A_3) = \frac{1}{4}$.
So $\forall i \neq j \quad P(A_i \cap A_j) = P(A_i)P(A_j)$, hence pairwise independent,
but $P(A_1 \cap A_2 \cap A_3) \neq P(A_1)P(A_2)P(A_3)$, hence not mutually independent.

■

Example 33: For n mutually independent events $A_1, ..., A_n$ show that

$$P(A_1 + \cdots + A_n) = 1 - (1 - P(A_1))(1 - P(A_2)) \cdots (1 - P(A_n)).$$

[3]
(2007.1dii)
SOLUTION:

(Induction on n):
Basis: For $n = 1$, LHS $= P(A_1) =$ RHS.
For $n = 2$,

$$\begin{aligned} \text{LHS} &= P(A_1 \cup A_2) \\ &= P(A_1) + P(A_2) - P(A_1 \cap A_2) \quad [\text{ by inclusion-exclusion principle}] \\ &= P(A_1) + P(A_2) - P(A_1)P(A_2) \quad [\because \text{ mutually independent}] \\ &= 1 - (1 - P(A_1))(1 - P(A_2)) = \text{RHS}, \end{aligned}$$

as required.
Hypothesis: We assume the result for $n = 1, ..., N$ for some $N \geq 2$.
Step: Shall show for $n = N + 1$.

$$\begin{aligned} \text{LHS} &= P(\underbrace{A_1 \cup \cdots \cup A_N}_{B_1} \cup \underbrace{A_{N+1}}_{B_2}) \\ &= 1 - (1 - P(B_1))(1 - P(B_2)) \quad [\text{ by the ``}n = 2\text{'' case}] \\ &= 1 - (1 - P(A_1))(1 - P(A_2)) \cdots (1 - P(A_N) \times (1 - P(B_2)) \quad [\text{ by induction hyp.}] \\ &= 1 - (1 - P(A_1))(1 - P(A_2)) \cdots (1 - P(A_N) \times (1 - P(A_{N+1})) = \text{RHS}, \end{aligned}$$

as required, completing the proof by the principle of mathematical induction.

■

Chapter II
Probability distributions

DAY 6
Random variable এবং probability distribution

এই অধ্যায়ে আমরা probability distribution নামে এমন একটা জিনিস শিখব যার গুরুত্ব probability এবং statistics-এর জগতে অপরিসীম। তার জন্য প্রথমে বুঝতে হবে random variable বলে আরেকটা জিনিসের কথা। একটা উদাহরণ নিয়ে শুরু করি। Probability theory-র অন্যান্য নানা জিনিসের মত এটারও জন্ম জুয়ার আসর থেকে। এই ধরণের খেলা পথেঘাটে বা মেলায় অনেক সময়ে খেলতে দেখা যায়।

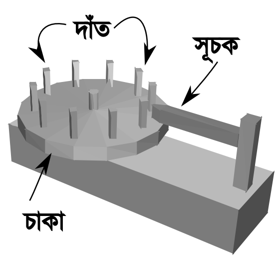
Example 1: খেলাটা হল এইরকম--একটা চাকা আছে, তার গায় অনেকগুলো দাঁত বসানো। Fig 1 দ্যাখো। ছবিতে দশটা দাঁত-ওয়ালা চাকা দেখানো হয়েছে। প্রতিটা দাঁতের পাশে একটা করে সংখ্যা লেখা আছে (সংখ্যাগুলো ছবিতে দেখাই নি)। চাকার বাইরে একটা কাঠি রয়েছে দ্যাখো, সেটা সূচকের কাজ করে। খেলার শুরুতে চাকাটাকে আচ্ছা করে জোরে ঘুরিয়ে দেওয়া হয়, খানিকক্ষণ ঘোরার পর যখন অবশেষে থামে, তখন সূচকের সামনে কোন দাঁতটা রয়েছে সেটা দেখা হয়। তার গায়ে যে সংখ্যাটা আছে তত টাকা তুমি পাবে।
এইখানে "তুমি কত টাকা পাবে" সেটা হল একটা random variable. এটা কেন variable সেটা তো বুঝতেই পারছ, কারণ লাভের অংকটা পরিবর্তনশীল। সেই পরিবর্তনটা নির্ভর করছে ভাগ্যের উপর, তাই random. ■

Random variable মানে এমন একটা variable যার value হবে কোনো real number[1], এবং সেটা নির্ধারিত হবে কোনো random experiment-এর outcome-এর উপর নির্ভর করে। গালভরা অংকের ভাষায় বললে একটা random variable মানে হল একটা random experiment-এর sample space থেকে $\mathbb{R}$-এ কোনো একটা function. যেমন আমাদের উদাহরণের চাকায় 10-টা দাঁত আছে, তাই random experiment-টার sample space হল

$$S = \{1 \text{ নং দাঁত}, 2 \text{ নং দাঁত}, \ldots, 10 \text{ নং দাঁত}\}.$$

[1] এমন random variable-ও সম্ভব যাদের value হয়তো real number নয় (complex number বা অন্যকিছু)। আমরা এ বইতে তাদের নিয়ে বড় একটা মাথা ঘামাব না।

Fig 1



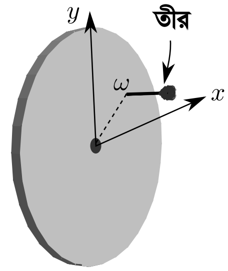


Fig 2

আর তোমার পাওয়া টাকার পরিমাণ হবে

$$\forall \omega \in S \quad X(w) = \text{“}w\text{-র পাশে লেখা সংখ্যা।”}$$

যেখানে ω হল random experiment-টার outcome. এই ω হল একটা গ্রীক অক্ষর, ছোটোহাতের ওমেগা। কেন লোকে random experiment-এর outcome-কে এই অক্ষর দিয়ে প্রকাশ করতে ভালোবাসে জানি না, কিন্তু সেটাই প্রচলিত নিয়ম, যদিও তুমি অন্য কোনো চিহ্ন ব্যবহার করলেও কোনো মহাভারত অশুদ্ধ হবে না। তবে বেশীর ভাগ সময়েই $X(\omega)$ না লিখে খালি X লেখা হয়।
এবার আরেকটা random variable-এর উদাহরণ দেখি। এটাও এক ধরণের খেলার সঙ্গে জড়িত যেটা অনেক সময়ে মেলায় লোকজনকে খেলতে দেখা যায়।

Example 2: একটা গোলাকার টার্গেট টাঙানো আছে (Fig 2), তার কেন্দ্রবিন্দুতে একটা ফুট্‌কি। কিছুটা দূর থেকে তীর ছুঁড়ে ওই টার্গেটে লাগাতে হবে। একেবারে ওই ফুট্‌কিতে লাগাতে পারলে তো খুবই ভালো, তা না হলেও প্রাইজ আছে, সেটা নির্ভর করবে ফুট্‌কি থেকে তীরের দূরত্বের উপর। এখানে এই দূরত্বটা হল একটা random variable. তীর ছোঁড়াটা হল random experiment, তার sample space-টা হল গোলাকার টার্গেটটাই। কোথায় তীরটা বিঁধেছে সেটাই হল এখানে outcome, যাকে যথারীতি আমরা ω নাম দেব। অর্থাৎ ω হল এখানে টার্গেটের উপর একটা random বিন্দু। কেন্দ্রের ফুট্‌কি থেকে এই ω-র দূরত্বটা নিয়ে আমরা মাথা ঘামাচ্ছি। এর নাম দিলাম $R(\omega)$, যেটা ছবিতে ড্যাশ্‌ড্যাশ্‌ লাইনটার দৈর্ঘ্য। এখানে $R(\omega)$ হল একটা random variable. ■

এই random variable-টার সঙ্গে আগেরটার পার্থক্য লক্ষ কর। আগেরটা খালি finite-সংখ্যক value নিতে পারত (যেহেতু দাঁতের সংখ্যা ছিল finite), কিন্তু এটা নিতে পারে $[0, r)$-এর মধ্যে যে কোনো value, যেখানে r হল টার্গেটটার radius (ব্যাসার্ধ)।
এবার আরেকটা উদাহরণ দেখি যেখানে আরেক রকমের random variable-এর সঙ্গে আমাদের পরিচয় হবে।

Example 3: একটা কয়েন নিয়ে টস করতে থাকো যতক্ষণ না head আসে। যেই প্রথম head আসবে অমনি থেমে যাও। তার মানে, যদি প্রথমেই head পড়ে অমনি থেমে যাবে। যদি tail পড়ে, তবে ফের টস করবে, যদি আবারও tail পড়ে তাহলে টস চালিয়ে যাবে। এইভাবে চলতে চলতে প্রথম head অব্‌ধি চলবে। এবার random variable-টা হল মোট কতবার টস করতে হয়েছে, সেই সংখ্যাটা। এখানে sample space-টা হবে

$$S = \{H,\ TH,\ TTH,\ TTTH,\ \cdots\}.$$

সুতরাং এখানে outcome (যাকে আমরা ω নাম দেব) হল এই sample space-এর কোনো একটা random সদস্য। যেমন হয়তো কখনো প্রথম দুবার tail পড়ে তৃতীয়বারে head পড়ল, সেক্ষেত্রে $\omega = TTH$, এইরকম। আমরা আমাদের random variable-টাকে লিখতে পারি

$$\forall \omega \in S \quad X(\omega) = \text{“}\omega\text{-র দৈর্ঘ্য।”}$$

এখানে $X(\omega)$ কী কী value নিতে পারে? উত্তর হল-- $1, 2, 3, ...$ ইত্যাদি যেকোনো positive integer (ধনাত্মক পূর্ণসংখ্যা)।
■

আমরা তিন ধরণের random variable দেখলাম। প্রথম ক্ষেত্রে random variable-টা খালি finite-সংখ্যক value নিতে পারত, দ্বিতীয় ক্ষেত্রে $[0, r)$-এর মধ্যে যেকোনো value. আর তৃতীয় ক্ষেত্রে countably infinite-সংখ্যক value. এর বাইরেও অন্যান্য ধরণের random variable সম্ভব, কিন্তু আমরা এই বইতে এই তিনটে ধরণ নিয়েই আলোচনা করব, কারণ বাস্তবে এদেরই প্রয়োগ সবচেয়ে বেশী। এই তিন ধরণের ক্ষেত্রেই statistical regularity লক্ষ করা যায়, মানে বার বার করে random experiment-টা করলে random ব্যাপারটা ক্রমশঃ মিলিয়ে গিয়ে একটা deterministic-এর মত regular আচরণ বেরিয়ে আসে। এবং এ থেকেই probability distribution-এর ধারণাটার জন্ম, যেটা এবার আলোচনা করব।

6.1 Probability distribution

কী করতে চলেছি বুঝে নিই। একটা random variable আছে X. তার মানে এর পিছনে একটা random experiment-ও আছে। সেই random experiment-টা যদি বার বার করা হয়, তবে X-এর বিভিন্ন value আসবে। আমরা জানতে চাই এই random variable-এর আচরণের মধ্যে statistically regular অংশটা কী। অর্থাৎ যতই বিভিন্ন value নিক না কেন, ওর কী ধর্ম আছে যেটা অপরিবর্তিত থাকছে? এই প্রশ্নটা কেন গুরুত্বপূর্ণ সেটা একটা উদাহরণ দেখলেই বুঝবে।

Example 4: ধরো তুমি কোনো বোর্ডের পরীক্ষা দেবে। এই বছর কী কী প্রশ্ন আসতে চলেছে তা তোমার অজানা। সেই অর্থে এটা তোমার কাছে একটা random experiment. বলাই বাহুল্য যে, তোমার আগ্রহ থাকবে পরীক্ষায় বসার আগেই প্রশ্নের প্যাটার্ণটা যথাসম্ভব জানতে পারা নিয়ে। সেই কাজে তোমায় সাহায্য করবে গত দশ বছরের প্রশ্নপত্র। সেগুলো হল সেই একই random experiment-এর গত দশখানা trial. অবশ্যই এমন কোনো স্থিরতা নেই যে, ঠিক ওই প্রশ্নগুলোই আবারও আসবে, কিন্তু সেখান থেকে তুমি প্রশ্নের প্যাটার্ণটা জানতে পারবে। এই প্যাটার্ণটা যাবতীয় randomness সত্ত্বেও অপরিবর্তিত থাকে। ■

বস্তুতঃ এই "অপরিবর্তিত" অংশটাই অতীতের সঙ্গে ভবিষ্যতের যোগসূত্র রচনা করে। সেই কারণেই আমরা অতীত ইতিহাস পর্যালোচনা করে ভবিষ্যতের কর্মপদ্ধতি ঠিক করতে পারি। যদি কোনো বছর সিলেবাসের বড় রকম পরিবর্তন হয়, সে বছর আর লোকে গত দশ বছরের প্রশ্ন নিয়ে বেশী আগ্রহী হয় না। আমাদের চারপাশে অহরহ যে নানা জিনিস হয়ে চলেছে তাদের অনেক কিছুরই কার্যকারণ আমরা সম্পূর্ণ বুঝি না। ঠিক যেন প্রকৃতি নানা random experiment করে চলেছে। একই random experiment-এর বারবার trial করা হচ্ছে। যেমন ক্যান্সার কী করে শুরু হয়, সেটা মানুষের অনেকটাই অজানা। ক্যান্সারের একজন নতুন রুগী মানে সেই random experiment-এর একটা নতুন trial. যে সব trial-গুলো ইতিমধ্যে হয়ে গেছে (মানে অতীতের রুগীরা), তাদেরকে খুঁটিয়ে দেখে আমরা random experiment-টার বিষয়ে জ্ঞান লাভ করছি। এবং সেই জ্ঞান কাজে লাগবে ভবিষ্যতের রুগীদের জন্য। যেমন ধরো একজন নাস্তিক ভদ্রলোক বেজায় ধূমপান করেন, তাঁর ফুসফুসে ক্যান্সার ধরা পড়ল। সেটা কী ভগবানে বিশ্বাস না করার শাস্তি, নাকি ধূমপানের ফল? খালি একজন রুগীর ভিত্তিতে সিদ্ধান্ত করা যাবে না। কিন্তু যখন বহু রুগী আসবে তখন দেখা যাবে যে তাদের মধ্যে বহু আস্তিকও আছেন, কিন্তু অধিকাংশই ধূমপান করতেন। তখন বোঝা যাবে যে, আস্তিক-নাস্তিক ব্যাপারটার এখানে কোনো গুরুত্ব নেই, ওরা নিতান্তই random, যেটা জোর দিয়ে বলা যায় সেটা হল ধূমপান করলে ফুসফুসে ক্যান্সার হবার সম্ভাবনা বেশী।

ঠিক একই ব্যাপার হয় যে কোনো random variable-এর ক্ষেত্রেই। ধরো X হল একটা random variable. তার পিছনে যে random experiment-টা আছে সেটাকে বার বার করতে করতে আমরা বুঝতে পারি X-এর বিভিন্ন value নেবার probability কত। গুছিয়ে বললে, যদি যেকোনো $A \subseteq \mathbb{R}$ দেওয়া থাকে তবে বলতে পারি $P(X \in A)$ কত। যেকোনো $A \subseteq \mathbb{R}$-র জন্য $P(X \in A)$ লেখাকেই আমরা বলি X-এর **probability distribution.**

যেকোনো $A \subseteq \mathbb{R}$-এর জন্যই $P(X \in A)$ লিখে ফেলা কম কথা নয়, কারণ $\mathbb{R}$-এর মধ্যে প্রচুর প্রচুর subset সম্ভব। তাদের সবার জন্য যদি probability-র তালিকা বানাতে হয় তবে সেই তালিকা কোনো দিনই ফুরোবে না। তাই আমরা খালি কিছু বিশেষ ধরণের subset-এর probability-ই লিখি। এই বিশেষ subset-গুলোকে এমনভাবে নির্বাচন করা হয়, যাতে বাকিদের probability-ও এদেরকে মিলিয়ে জুলিয়ে প্রকাশ করা যায়।

এই বিশেষ subset-গুলো বিভিন্ন সময়ে বিভিন্ন নেওয়া হয়, তার উপর নির্ভর করে তিনটে কায়দা আমরা এবার আলোচনা করব।

6.1.1 প্রথম কায়দা (PMF)

যদি কোনো random variable কেবল countable-সংখ্যক value নিতে পারে $\{x_1, x_2, ...\}$, তবে তাকে বলে একটা **discrete random variable.** যেমন দাঁতওয়ালা চাকা আর বারবার কয়েন টসের উদাহরণের বেলায়। এরকম ক্ষেত্রে প্রতিটা $\{x_i\}$-এর probability জানলেই বাকি সব probability বার করে দেওয়া যায়। তাই discrete random

variable-দের probability distribution-কে প্রকাশ করা যায় probability mass function (PMF) দিয়ে, যার সংজ্ঞাটা এইরকম--

> DEFINITION: **Probability mass function (PMF)**
>
> Let X be a random variable taking only countably many values. Then its probability mass function (PMF) is defined as the function $p:\mathbb{R}\to[0,1]$ where
>
> $$p(x) = P(X=x).$$

এবার একটা উদাহরণ দেখি যাতে বুঝবে যে একটা discrete random variable X-এর PMF দেওয়া থাকলে তা থেকে কী করে যেকোনো $A\subseteq\mathbb{R}$-এর জন্যই $P(X\in A)$ বার করা যায়।

Example 5: বলা আছে যে, X একটা discrete random variable, যার PMF হল

$$p(x) = \begin{cases} 2^{-x} & \text{if } x\in\{1,2,3,...\} \\ 0 & \text{otherwise} \end{cases}$$

তাহলে নীচের প্রতিটা ক্ষেত্রে $P(X\in A)$ বার কর--
(i) $A=\{1,2,3\}$. (ii) $A=[-5,8]$. (iii) $A=\{1,3,5,\cdots\}$. (iv) $A=\mathbb{R}$.
SOLUTION: লক্ষ কর যে এখানে X খালি $1,2,3,...$ ইত্যাদি value-ই নিতে পারে।
প্রথম ক্ষেত্রে তো খালি যোগ করে দিলেই হবে--

$$\begin{aligned} P(X\in\{1,2,3\}) &= P(\{X=1\}\cup\{X=2\}\cup\{X=3\}) = P(X=1)+P(X=2)+P(X=3) \\ &= p(1)+p(2)+p(3) = \frac{1}{2}+\frac{1}{4}+\frac{1}{8} = \frac{7}{8}. \end{aligned}$$

দ্বিতীয় ক্ষেত্রে দেখে নিতে হবে $[-5,8]$-এর মধ্যে কটা value আছে যেগুলো X সত্যিই নিতে পারে, মানে

$$[-5,8]\cap\{1,2,3,...\} = \{1,2,3,...,8\}.$$

সুতরাং $p(1)+\cdots+p(8)$ বার করলেই উত্তর পেয়ে যাবে।
তৃতীয় ক্ষেত্রেও একটা যোগ করতে হবে, তবে সেটা হবে একটা infinite series--

$$p(1)+p(3)+p(5)+\cdots = \tfrac{1}{2}+\left(\tfrac{1}{2}\right)^3+\left(\tfrac{1}{2}\right)^5+\cdots.$$

এটা একটা geometric series, যার প্রথম term হল $\frac{1}{2}$ আর common ratio হল $\frac{1}{4}$. তাই যোগফলটা হবে $\frac{1}{2}\times\frac{1}{1-1/4}=\frac{2}{3}$.
চতুর্থ ক্ষেত্রেও একটা geometric series আসবে, এখানে সবকিছু যোগ হবে--

$$p(1)+p(2)+p(3)+\cdots = \tfrac{1}{2}+\left(\tfrac{1}{2}\right)^2+\left(\tfrac{1}{2}\right)^3+\cdots = \cdots = 1.$$

এখানে যে উত্তর 1 হবে, সেটা কিন্তু এমনিতেই বোঝা যাচ্ছিল। এখানে আমরা সব কিছু যোগ করেছি, কোনো কিছুই বাদ দিই নি। তার মানে পুরো sample space-এর probability-টাই বেরোবে। Axiom থেকেই জানি যে, পুরো sample space-এর probability সবসময়ে 1 হয়। সুতরাং যদি উত্তর 1 না আসত, তবে বুঝতে হত যে, PMF-টায় কিছু গলদ

আছে। ■

এবার একটা কায়দা বলি যেটা দিয়ে কোনো $p : \mathbb{R} \to [0,1]$ দেওয়া থাকলেই তুমি চট করে বুঝে যেতে পারবে সেটা একটা PMF কিনা।

THEOREM

A function $p : \mathbb{R} \to [0,1]$ is a PMF if and only if it satisfies all these conditions:

1. there is a countable set $S \subseteq \mathbb{R}$ such that $\forall x \notin S \quad p(x) = 0$.
2. $\forall x \in S \quad p(x) > 0$.
3. $\sum_{x \in S} p(x) = 1$.

একটা PMF-কে যে এই তিনটে শর্ত পালন করতেই হবে সেটা বোঝা কঠিন নয়। সেটা probability axiom-গুলো থেকেই চলে আসে। কিন্তু এই theorem-টার কঠিন অংশটা হল উল্টো দিকটা--এই তিনটে শর্ত পালন করলেই এটা একটা PMF হতে বাধ্য, মানে কোনো $p : \mathbb{R} \to [0,1]$ যদি এই শর্তগুলো পালন করে, তবে এমন random variable থাকতে বাধ্য (ধরো X) যাতে $\forall x \in \mathbb{R} \quad P(X = x) = p(x)$ হয়। আমরা তার প্রমাণের মধ্যে এ বইতে যাব না।

6.1.2 দ্বিতীয় কায়দা (PDF)

সব random variable-ই অবশ্য discrete হয় না, যেমন আমাদের তীর ছোঁড়ার উদাহরণের random variable-টা। যারা uncountably infinite-সংখ্যক value নেয়, তাদের বেলায় PMF-এর কায়দাটা কাজ করে না। এদের অনেকের ক্ষেত্রে probability density function (PDF) বলে অন্য একটা জিনিস কাজ করে, যেটা আমরা এবার শিখব। এখানেও মূল ব্যাপারটা একই, আমরা বিশেষ ধরণের কিছু $A \subseteq \mathbb{R}$-এর জন্য $P(X \in A)$ বার করার ফর্মুলা দেব, তাদের মিলিয়ে মিশিয়ে বাকি সব probability-ও বার করা যাবে। এখানে এই বিশেষ ধরণের A-রা হল closed, bounded interval-রা, মানে $[a,b]$ জাতীয় set. এই কায়দাটা একটু জটিল তাই প্রথমেই সংজ্ঞাটা দিয়ে নিই, তারপর বোঝাচ্ছি।

Example 6: Define the probability density function of a random variable. [2] **(2014.2a)**

SOLUTION:

DEFINITION: Probability density function (PDF)

A random variable X is said to have a <u>probability density function (PDF)</u> $f : \mathbb{R} \to [0,\infty)$ if

$$\forall a \le b \in \mathbb{R} \quad P(X \in [a,b]) = \int_a^b f(x)dx.$$

■

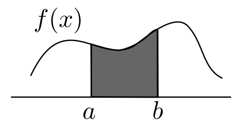


Fig 3

সংজ্ঞার প্যাঁচটা লক্ষ করো--pdf-টা কিন্তু সরাসরি probability দিচ্ছে না, ওটাকে integrate করলে তবে probability বেরোচ্ছে। Fig 3 দ্যাখো। এখানে বলে রাখি যে, PDF-কে অনেক সময়ে খালি density-ও বলা হয়। এবার দেখা যাক pdf জানা থাকলে কী করে বিভিন্ন ধরণের $A \subseteq \mathbb{R}$-এর জন্য $P(X \in A)$ বার করতে হয়।

Example 7: ধরো X একটা random variable, যার density হল

$$f(x) = \begin{cases} e^{-x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

তবে নীচের প্রতি ক্ষেত্রে $P(X \in A)$ বার করো।
(*i*) $A = [2,3]$. (*ii*) $A = \{3\}$. (*iii*) $A = \{2\}$. (*iv*) $A = [2,3)$. (*v*) $A = (2,3)$.
SOLUTION:

(i) $[2,3]$ একটা closed, bounded interval-ই আছে, তাই সরাসরি সংজ্ঞা লাগানো যাবে--

$$P(X \in [2,3]) = \int_2^3 f(x)dx = \int_2^3 e^{-x}dx = e^{-2} - e^{-3}.$$

(ii) 3-ও কিন্তু আসলে একটা closed, bounded interval, কারণ $\{3\} = [3,3]$. তাই এখানেও সরাসরি সংজ্ঞা লাগানো যাবে--

$$P(X \in \{3\}) = P(X \in [3,3]) = \int_3^3 f(x)dx = 0.$$

এখানে আর integration-টা কষ্ট করে করতেও হল না, কারণ দুই প্রান্তেই যেহেতু একই সংখ্যা তাই উত্তর শূন্য হতে বাধ্য!

(iii) এখানেও একই যুক্তিতে উত্তর হবে 0.

(iv) এখানে লক্ষ করো যে, $[2,3] = [2,3) \cup \{3\}$.
যেহেতু $[2,3) \cap \{3\} = \phi$, তাই $P(X \in [2,3]) = P(X \in [2,3)) + P(X \in \{3\})$.
মানে $P(X \in [2,3)) = P(X \in [2,3]) - P(X \in \{3\})$. আমরা এক্ষুণি দেখলাম যে এর মধ্যে প্রথমটা হল $e^{-2} - e^{-3}$, আর দ্বিতীয়টা 0. সুতরাং এখানেও উত্তর $e^{-2} - e^{-3}$.

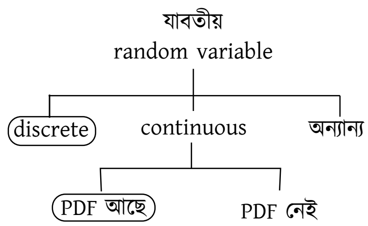
(v) এখানেও একই যুক্তিতে উত্তর হবে $e^{-2} - e^{-3}$.

■

সুতরাং লক্ষ কর যে, X-এর যদি একটা pdf থাকে, তবে--

- $P(X \in [a,b]) = P(X \in (a,b]) = P(X \in [a,b)) = P(X \in (a,b))$ হতে বাধ্য।
- যে কোনো singleton set-এর probability হবে 0.

এই প্রসঙ্গে বলে রাখি যে, যেসব random variable-দের বেলায় সব singleton set-এর probability হয় 0, তাদের বলে **continuous random variable.** সুতরাং আমরা শিখলাম যে, যে সব random variable-এর pdf থাকে তারা continuous হয়। এর উল্টো দিকটা কিন্তু ঠিক নয়, continuous random variable হলেই যে তার pdf থাকবে এমন কোনো কথা নেই, যদিও pdf-বিহীন continuous random variable-এর উদাহরণ বানানো মোটেই সহজ নয়। এই প্রসঙ্গে random variable-দের একটা শ্রেণীবিভাগ শিখে রাখি--



এর মধ্যে এ বইতে আমরা খালি discrete-দের নিয়ে আর যাদের PDF আছে তাদের নিয়েই আলোচনা করব। এদেরকে গোল করে ঘিরে দিয়েছি। তবে জানা থাকা ভালো যে, এমন random variable-ও সম্ভব যারা discrete-ও নয়, continuous-ও নয়। শ্রেণীবিভাগটা থেকেই দেখতে পাচ্ছ যে একটা random variable যদি continuous হয়, তাহলে তার PDF নাও থাকতে পারে। কিন্তু PDF-বিহীন continuous random variable-দের নিয়ে আমরা মাথা ঘামাব না। তাই এ বইতে যদি কোনো প্রশ্নে দ্যাখো যে, একটা random variable-কে continuous বলে দিয়েছে, তবে ধরে নিতে পারো যে, তার PDF আছে।

Pdf ব্যবহার করে আরো probability বার করার উদাহরণ দেখা যাক।

Example 8: আবার সেই আগের অংকের density নিয়েই কাজ করব, মানে

$$f(x) = \begin{cases} e^{-x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

নীচের প্রতি ক্ষেত্রে $P(X \in A)$ বার করো।

(*i*) $A = [-2, 2]$ (*ii*) $A = [1, \infty)$ (*iii*) $A = (-\infty, 2]$ (*iv*) $A = (-\infty, \infty)$

SOLUTION:

(i) এখানে সরাসরি সংজ্ঞা লাগানো যাবে--

$$P(X \in [-2, 2]) = \int_{-2}^{2} f(x)dx = \int_{-2}^{0} 0dx + \int_{0}^{2} e^{-x}dx = 1 - e^{-2}.$$

কীভাবে দুভাগে ভেঙে নিতে হল দেখলে? এখানে এই প্যাঁচটাই খালি শেখার।

(ii) এখানে একটা কৌশল লাগবে। আমরা প্রথম অধ্যায়ে শিখেছিলাম যে, যদি $\{A_n\}_n$ একটা increasing sequence of events হয় আর $A = \cup_n A_n$ হয়, তবে $P(A_n) \to P(A)$ হবেই! সেইটাই এবার কাজে লাগবে। যদি $A_n = [1, n]$ নিই, তবে লক্ষ কর যে, A_n-গুলো বাড়তে বাড়তে $[1, \infty)$-তে যায়। সুতরাং $P(A_n) \to P(A)$ হবে। মানে

$$P(X \in [1, \infty)) = \lim_{n \to \infty} \int_{1}^{n} f(x)dx = \int_{1}^{\infty} f(x)dx = \cdots = e^{-1}.$$

(iii) আগের যুক্তিটা যদি বুঝে থাকো তবে এখানেও অসুবিধা হবে না--

$$P(X \in (-\infty, 2]) = \int_{-\infty}^{2} f(x)dx = \int_{0}^{2} e^{-x}dx = \cdots = 1 - e^{-2}.$$

(iv) এখানে যে উত্তরটা 1 হবেই সেটা নিশ্চয়ই বুঝতে পারছ। কারণ এখানে গোটা $\mathbb{R}$-এর উপরে integration হবে, মানে পুরো sample space-এর probability, যেটা 1 হতে বাধ্য। Integration-টা করে মিলিয়ে নিও যে সত্যিই 1 আসছে।

■

তার মানে সব ধরণের interval-এর বেলাতেই integration-এর কায়দাটা কাজ করে, $[a, b]$ বা (a, b) বা (a, ∞) বা $(-\infty, b)$, যাই হোক না কেন।
যদি তোমাকে কোনো একটা function দেওয়া হয় $f : \mathbb{R} \to \mathbb{R}$, এবং দিয়ে জানতে চাওয়া হয় এটা কোনো random variable-এর PDF হতে পারে কিনা, তাহলে নীচের theorem-টা খুব কাজে দেবে--

THEOREM

A function $f : \mathbb{R} \to \mathbb{R}$ is PDF of some random variable if and only if satisfies the following conditions:

1. $\forall x \in \mathbb{R} \quad f(x) \geq 0$.
2. $\int_{-\infty}^{\infty} f(x) dx = 1$.

6.1.3 তৃতীয় কায়দা (CDF)

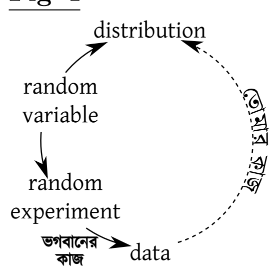
আমরা দেখেছি যে, সব discrete random variable-এর বেলাতেই PMF পাওয়া যায়। যে random variable-রা discrete নয়, তাদের কারুর কারুর ক্ষেত্রে pdf থাকে, সকলের ক্ষেত্রে নয়। প্রশ্ন হল বাকি random variable-এর distribution কী করে প্রকাশ করা যাবে? তার জন্য একটা general কায়দা আছে যেটা যে কোনো random variable-এর ক্ষেত্রেই খাটে। কায়দাটার নাম হল cumulative distribution function, যার সংজ্ঞা এইরকম--

DEFINITION: **Cumulative distribution function (CDF)**

For any random variable X its cumulative distribution function (CDF) is defined as the function $F : \mathbb{R} \to [0, 1]$ where

$$\forall x \in \mathbb{R} \quad F(x) = P(X \leq x).$$

Fig 4



DAY 7
Statistics-এর প্রাণ-ভোমরা

এবার আমরা কয়েকটা জিনিস আলোচনা করব যেগুলো না বোঝা থাকলে statistics-এর অধিকাংশ অংকই নিরর্থক মনে হবে। জিনিসগুলো এক দিক দিয়ে দেখলে হয়তো নিতান্তই সহজ, এতই সহজ যে বলাই বাহুল্য। এবং সেই কারণেই বোধহয় অনেক বইতেই কথাগুলো বিশদ করে বলা থাকে না। কিন্তু বহু ছাত্রকেই দেখেছি এই জিনিস কয়টা না বোঝার ফলে খাবি খাচ্ছে। সেই কারণেই আজকের দিনটা আমরা এই "সহজ" জিনিসগুলোর আলোচনা করে কাটাব।

7.1 Statistical set up

এ বইতে আমরা যা যা statistics করব, তার সবেরই মূল set up-টা একইরকম--

- একটা random variable থাকবে যাকে নিয়ে আমরা মাথা ঘামাচ্ছি। Random variable, অর্থাৎ কিনা এমন কোনো variable যার পরিবর্তনশীলতার সম্পূর্ণ বৈজ্ঞানিক ব্যাখ্যা আমাদের জানা নেই (যেমন IQ, একজনের বেশী আরেকজনের কম কেন হয়, তার কোনো সদুত্তর আজও জানা নেই)।

- আগেই বলেছি যে একটা random variable-এর প্রতিটা ওঠানামা নিয়ে দুশ্চিন্তা করার মানে হয় না, গুরুত্বপূর্ণ খালি তার মধ্যে যেটুকু statistically regular অংশ সেটুকুই, মানে তার probability distribution.

- কিন্তু probability distribution-টা সরাসরি জেনে ফেলবার কোনো কায়দা নেই। তাই আমাদের দরকার data. সেটা আসে এইভাবে--যে কোনো random variable-এর পিছনেই একটা random experiment থাকে, যেটাকে বারবার independent-ভাবে করা যায়। এইভাবে n-বার করলে আমরা X-এর n-খানা value পাব, যাদের বলে X-এর একেকটা **realisation.** সাধারণতঃ আমরা এদেরকে $X_1, ..., X_n$ হিসেবে লিখি।

এইভাবে "বার বার করে একই random experiment করা"--সেটা আমাদের চারপাশে অহরহ হয়ে চলেছে (যেমন দশটা ছাত্রের IQ মাপা মানেই IQ নামক random variable-এর দশটা realisation পাওয়া)। ভাবতে পারো যেন random experiment-টা ভগবান (কিংবা প্রকৃতি) বারংবার করেই চলেছেন, আমরা খালি তার ফলাফলটুকু দেখে নিচ্ছি। এরাই হল আমাদের data. এবার statistician হিসেবে তোমার কাজ হল data-র ভিত্তিতে distribution-টার বিষয়ে জ্ঞান আহরণ করা। এই কাজটাকে বলে statistical analysis. Fig 4 দেখলে সুবিধা হবে।
এই বর্ণনাটা একটু কঠিন লাগতে পারে। নীচের উদাহরণটা দেখলে সুবিধা হবে।

Example 9: একজন তোমায় একটা কয়েন দেখিয়ে বলল যে, সেটায় head পড়লে তোমাকে সে 500 টাকা দেবে,

আর tail পড়লে তুমি তাকে 500 টাকা দেবে। এই খেলাটা তোমার পক্ষে লাভজনক হবে কিনা, সেটা নির্ভর করছে কয়েনে কী আসবে তার উপর। সেটাই হল আমাদের random variable. কিন্তু ঠিক কী পড়বে তার কোনো স্থিরতা নেই। যে জিনিসটার স্থিরতা আছে, সেটা হল head আর tail পড়ার probability. এদের নিয়েই আমাদের random variable-টার probability distribution. সেই distribution-টা তোমার জানা নেই, সেটা তোমায় আন্দাজ করতে হবে। তার জন্য তুমি চাইবে কয়েনটা আগেভাগে কয়েকবার টস করে দেখে নিতে। ধরো 100-বার টস করলে। তাতে head এল মোটে 11 বার, আর tail এল অবশিষ্ট 89 বার। এই হল তোমার data. এটা থেকে তুমি আন্দাজ করবে যে, কয়েনটা গোলমেলে, কারণ এখানে $P(\text{head})$ মোটেই $\frac{1}{2}$ নয়, বরং 11/100-র ধারেকাছে কিছু একটা। এই সিদ্ধান্তটাই হল এখানে statistical analysis. ■

এইখানে কয়েকটা নতুন ভাষা শিখে রাখা দরকার--

- $X_1, ..., X_n$ সকলেই একই random experiment-কে বার বার independent-ভাবে করে পাওয়া। এই কথাটাকে আমরা সংক্ষেপে বলি "$X_1, ..., X_n$ are IID" এখানে IID মানে "independently and identically distributed".

- একই কথা আরেকভাবেও বলা হয়ে থাকে--"$X_1, ..., X_n$ are a **random sample** from a distribution.

- Sample শব্দটার এমনি মানে হল "নমুনা", যেমন মেডিকেল রিপ্রেজেনটেটিভরা এসে ডাক্তারবাবুদের ওষুধের sample দিয়ে যায়। এখন, নমুনার কথা যখনই ওঠে, তখন স্বভাবতঃই প্রশ্ন আসে-- কীসের নমুনা, এরা কাদের প্রতিনিধি? যেসব জিনিসের এরা প্রতিনিধিত্ব করছে, তাদের পুরো set-টাকে বলা হয় **population.** যেমন, ওষুধের বেলায় কোনো কোম্পানি যা যা ওষুধ তৈরী করেছে, তাদের set-টাই হল population. কিন্তু আমরা যখন বলি $X_1,...,X_n$ একটা random sample, তখন সেটা ঠিক ওষুধের sample-এর মত ব্যাপার নয়। কিন্তু তাও অনেকে এখানেও একটা population কল্পনা করে নিতে ভালোবাসে। ধরো পাঁচবার একটা কয়েন টস করে পেলে $X_1,...,X_5$. এখানে population কী কল্পনা করব? উত্তর হল "যতবার কয়েনটা টস করা যায়" তার set-টাই হল population. বলাই বাহুল্য যে, একটা কয়েনকে তুমি যতবার খুশি টস করতে পারো[2]। এইরকম কাল্পনিক population-কে আমরা **infinite population** বলি। তাই

  $X_1,...,X_n$ are IID with a distribution

  বলা আর

  $X_1,...,X_n$ are a random sample from an (infinite) population

  বলা একই কথা। তবে এই দ্বিতীয় অর্থে population শব্দের ব্যবহার বর্তমান কালে কমে এসেছে। কোনো আন্তর্জাতিক মানের আধুনিক বইতেই সাধারণতঃ এই ভাষা আর ব্যবহার করা হয় না। তবে প্রথম অর্থে (মানে finite ক্ষেত্রে) population শব্দটা এখনও বহুল ব্যবহৃত। যদিও এই বইতে প্রথম অর্থটার প্রায় কোনোই প্রয়োজন হবে না।

Example 10: Explain the terms: population, random sample.[2] **(2005)**

> Population: When we want to draw inference about the elements of a (usually very large) set, that set is called a population. Such a set is typically too large to be handled directly.
>
> For example, if we want to study the income of Indians, then the set of all Indians is the population.
>
> The term infinite population is used to mean a probability distribution.
>
> Random sample: By a random sample from a (finite) population, we mean a subset (possibly with repetition) of the population that is selected based on some random mechanism.
>
> A random sample from an (infinite) population means a number of random variables that are independently and identically distributed according to the probability distribution of the population.

■

Exercise 1: Explain the terms (i) population, (ii) sample.[2] **(2008,2013)**
HINT:
যদিও এখানে random sample না বলে খালি sample বলেছে, তাও এটা আসলে আগের প্রশ্নটাই। ■

[2] হাত ব্যথা না হলে, আর টস করতে করতে কয়েনটা ক্ষয়ে না গেলে।

Exercise 2: Define a sample.[2] **(2006.4a)**
HINT: আগের অংকেই এর উত্তর আছে। ■

7.2 Data থেকে distribution বার করা

এবার আমরা দেখব data থেকে distribution কী করে বার করতে হয় (বা বলা ভালো আন্দাজ করতে হয়)। আমাদের কাছে একটা random variable আছে, X, যার distribution-টা আমরা জানতে চাই। Data হিসেবে আছে একটা random sample, মানে X-এর কিছু realisation, $X_1, ..., X_n$, যারা IID.
এদের ব্যবহার করে X-এর distribution আন্দাজ করার জন্য আমরা তিনটে কায়দা শিখব। কোন কায়দাটা লাগবে সেটা নির্ভর করবে X-এর distribution-টার কোন রূপটা আমরা চাই, PMF, নাকি density, নাকি CDF, তার উপর।
তিনটে কায়দারই মূলে আছে সেই একই কথা, যেটা আমরা আগেই শিখেছি, এবং এখানে আরেকবার মনে করিয়ে দিই--

> ধরো একটা random experiment আছে, আর A হল তার কোনো একটা event. তবে বার বার করে independent-ভাবে experiment-টাকে করে দ্যাখো A ঘটার relative frequency কত। মানে n-বার করার পর
>
> $$\frac{A \text{ কতবার ঘটেছে}}{n}$$
>
> বার কর। তবে সেই সংখ্যাগুলো $P(A)$-র দিকে এগিয়ে যাবে।

এইবার আমরা তিনটে কায়দা একে একে বলব।

7.2.1 PMF আন্দাজ করার কায়দা

এখানে আমরা ধরে নেব যে X-এর একটা PMF আছে (যেটা অজানা)। যেহেতু $X_1, ..., X_n$ হল X-এরই কিছু realisation, তাই ওদেরও PMF সেটাই। এই PMF-টাকেই আমরা আন্দাজ করতে চাই।

Example 11: আমরা আবার সেই দশ দাঁতওয়ালা চাকার উদাহরণে ফিরে যাই। ধর, যত নম্বর দাঁত তার পাশে সেই সংখ্যাটাই লেখা। মানে, প্রথম দাঁতের পাশে 1 লেখা, দ্বিতীয় দাঁতের পাশে 2, এইরকম। এখানে statistical regularity কি করে আসছে সেটা বোঝার জন্য এই random experiment-টা বার বার করা যাক, ধরো 50-বার করলাম। এখানে যেন ভেবে বোসো না যে চাকাটা সব দিকেই সমান ভারী। জুয়াড়িদের চাকায় নানা কারিকুরি করা থাকে। সাদা চোখে ধরা যায় না, কিন্তু হয়তো একদিক সামান্য হেলানো আছে, ফলে সেই দিকের দাঁতগুলো বেশীবার করে আসে। বা হয়তো ওই সূচক কাঠিটার মধ্যে একটা ছোটো চুম্বক লুকোনো আছে, আর কয়েকটা দাঁতের মধ্যে হয়তো বিভিন্ন উচ্চতায় লোহার পেরেক আছে। তার উপর নির্ভর করে সেই দাঁতগুলো সূচকের কাছে এসে আটকে যাবার probability কম বেশী হবে। এরকম একটা জালি চাকায় 50 বার খেলার পর কতবার কত করে টাকা পেলাম দেখি--

X-এর value	1	2	3	4	5	6	7	8	9	10
Frequency	4	4	5	5	6	10	3	3	4	**6**
Relative frequency	0.08	0.08	0.1	0.1	0.12	0.2	0.06	0.06	0.08	0.12

এই টেবিলটা বুঝে নাও। আমি 50 বার খেলার পরে দেখছি যে 10 টাকা পেয়েছি ঠিক 6 বার। সেই জন্যেই দ্বিতীয় লাইনের শেষ সংখ্যাটা 6 (মোটা করে দেখিয়েছি)। আবার 7 টাকা পেয়েছি ঠিক 3 বার, 8 টাকাও পেয়েছি 3 বার। যেহেতু ঠিক 50-বার খেলার পরে এই টেবিলটা বানানো, তাই দ্বিতীয় লাইনের সব সংখ্যা যোগ করলে হবে 50. তৃতীয় লাইনে relative frequency হিসেব করেছি এইভাবে--10 টাকা জিতেছি 50 বার খেলার মধ্যে 6 বার, অর্থাৎ 10 টাকা জেতার relative frequency হল $\frac{6}{50} = 0.12$. এইভাবে তৃতীয় লাইনের সংখ্যাগুলো পাওয়া গেছে দ্বিতীয় লাইনের সংখ্যাগুলোকে 50 দিয়ে ভাগ করে।

Exercise 3: চট করে বলো তো সবগুলো relative frequency-র যোগফল কত হতে বাধ্য? সত্যি সত্যি সবগুলোকে যোগ করে দেখা দরকার নেই কিন্তু, এমনি চিন্তা করলেই উত্তরটা পেয়ে যাবে। ■

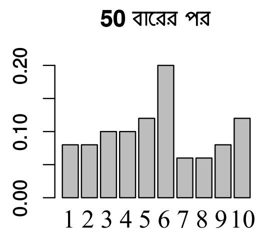


Fig 5

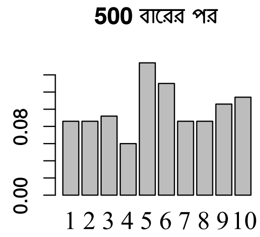


Fig 6

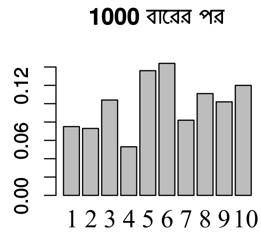


Fig 7

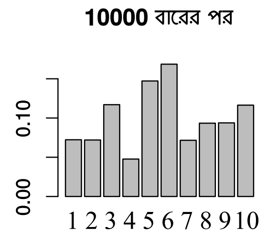


Fig 8

এই relative frequency-গুলোকে ছবিতে এঁকেছি Fig 5-এ। প্রত্যেকটা সংখ্যার জন্য একটা করে rectangle আঁকা হয়েছে। এদের সবারই প্রস্থ সমান। প্রস্থটা আসলে গুরুত্বপূর্ণ নয়, কাগজে ধরানোর জন্য যতটা সরু করতে হয় তাই করেছি। গুরুত্বপূর্ণ হল উচ্চতাটা, সেটাই হল তৃতীয় লাইনের সংখ্যাগুলোর সমান। যেমন, প্রথম দুটো rectangle-এর উচ্চতা হল 0.08, কারণ তৃতীয় লাইনের প্রথম দুটো সংখ্যাও তাই। এই যে ছবিটা পেলাম, একে বলে একটা **barplot.** এখানে plot মানে ছবি, আর ওই rectangle-গুলোকে একেকটা bar বলে।
Barplot দিয়ে কীভাবে PMF বার করা যায় সেটা দেখা যাক। মনে করো তুমি খালি প্রথম bar-টার উচ্চতা দেখছ। সেটা হল "$X = 1$" হওয়ার relative frequency, অর্থাৎ--

$$\frac{\text{কতবার } X = 1 \text{ হয়েছে}}{\text{মোট কতবার খেলা হয়েছে}}$$

সুতরাং প্রথম bar-টার উচ্চতা $P(X = 1)$-এর দিকে এগিয়ে যাবে। একইভাবে দ্বিতীয় bar-টার উচ্চতা যাবে $P(X = 2)$-তে, এইরকম। তার মানে যতই বেশীবার খেলবে ততই barplot-টা ক্রমশঃ PMF-এর ছবির দিকে এগিয়ে যাবে। খেলেই দেখা যাক--
ধরো আরো 450 বার একই খেলা খেললাম, মানে মোট $50 + 450 = 500$ বার খেলা হল। এর ভিত্তিতে আমরা আবার একটা টেবিল পাব। এখানেও আমরা relative frequency-গুলো বার করতে পারব (এবার ভাগ করতে হবে 500 দিয়ে)। তাহলে যে barplot-টা পেয়েছি সেটা দেখিয়েছি Fig 6-এ। লক্ষ কর ছবিটা অনেকটাই বদলে গেছে। সেটাই স্বাভাবিক, কারণ হাজার হোক খেলাটা তো random!
আরও বহুবার খেলতে থাকি। মোট 1000 বার খেলার পরে পেয়েছি Fig 7-এর ছবিটা। লক্ষ কর যে চেহারাটা আবারও বদলে গেছে, কিন্তু তাও খানিকটা আদল কিন্তু একই আছে। যদি আরও বহুবার খেলি, মনে করো 10000 বার, তবে পাবে Fig 8-এর মত একটা ছবি। এবার লক্ষ কর চেহারাটা অনেকটাই স্থির হয়ে এসেছে। অর্থাৎ statistical regularity-এর দরুণ convergence প্রকাশ পেতে শুরু করেছে। ■

এইভাবেই barplot ব্যবহার করে PMF আন্দাজ করা যায়।

7.2.2 pdfআন্দাজ করার কায়দা

এবার আমরা ধরে নেব যে X-এর একটা pdf আছে, যেটা আমরা জানি না। আমাদের হাতে যে random sample-টা আছে, মানে $X_1, ..., X_n$, তাদের ব্যবহার করে আমরা এই pdf-টাকে আন্দাজ করতে চাই।

Example 12: এবার আমরা তীর ছোঁড়ার উদাহরণটা আবার দেখব। ধরো টার্গেটটার radius হল $r = 5$. এখানে X

একটানা value নেয়, $[0, 5)$-এর মধ্যে যেকোনো value-ই নিতে পারে। আমরা প্রথমে $[0, 5)$-কে কয়েকটা subinterval-এ ভেঙে নেব, যেমন ধরো

$$[0, 5) = [0, 1) \cup [1, 2) \cup [2, 4) \cup [4, 5).$$

Statistics-এর ভাষায় এই subinterval-দের বলে একেকটা **class.** বস্তুতঃ অধিকাংশ টার্গেটে এরকম ভাগ করাই থাকে। যেমন দেখিয়েছি Fig 9-এ। এখানে সবচেয়ে ভিতরের গোলটার radius হল 1, তার পরের গোলটার radius হল 2, তারপরেরটার 4, আর পুরো টার্গেটটার radius তো 5 আছেই। লক্ষ কর যে, class-গুলোর প্রস্থ (যাকে বলে **class width**),

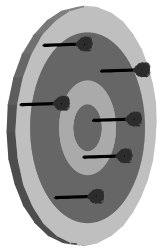


Fig 9

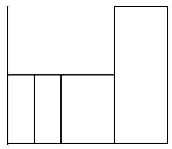


Fig 10

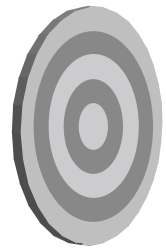


Fig 11

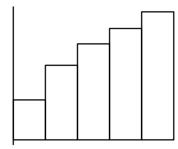


Fig 12

তারা অসমান হলেও আপত্তি নেই। এর ফলে টার্গেটটা কয়েকটা ভাগে ভেঙে গেছে--একদম ভিতরে একটা গোল, আর তাকে ঘিরে পর পর কয়েকটা রিং। Fig 9-এ দেখা যাচ্ছে যে, 6-বার তীর ছোঁড়া হয়েছে। কোথায় কটা বিঁধেছে তার ভিত্তিতে এই টেবিলটা পাচ্ছি--

class	$[0,1)$	$[1,2)$	$[2,4)$	$[4,5)$
frequency	1	1	2	2
relative frequency	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$
relative frequency density	$\frac{1/6}{1-0}$	$\frac{1/6}{2-1}$	$\frac{2/6}{4-2}$	$\frac{2/6}{5-4}$

এখানে প্রথম তিন লাইন আগের মতই। নতুন খালি চতুর্থ লাইনটা। এটার প্রয়োজন হয়েছে কারণ এখানে class width-গুলো সবাই সমান নয়। সেই তারতম্যটাকে হিসেবে নেওয়ার জন্য class width দিয়ে ভাগ করেছি। মানে তুমি এখন একক দৈর্ঘ্যে relative frequency-র পরিমাণ মাপছ। একক আয়তনের ভরকে যেমন ঘনত্ব বা density বলে, তেমনি একক দৈর্ঘ্যের relative frequency-কে বলে **relative frequency density.** কেন density নেওয়ার প্রয়োজন হচ্ছে সে তো বুঝতেই পারছ-- এক বস্তা তুলোর ওজন একটা ছোট্টো লোহার বলের চেয়ে বেশী হতেই পারে, তা বলে এমন বলা মোটেই উচিত হবে না যে, "তুলো লোহার থেকে ভারী"।

এইবার আবার আমরা barplot-এর মতই একটা ছবি আঁকব, যেটা দেখিয়েছি Fig 10-এ। এইরকম ছবিকে বলে **histogram.** আগে যে barplot এঁকেছিলাম তার সঙ্গে এর তিনটে গুরুত্বপূর্ণ পার্থক্য আছে--

1. এখানে rectangle-গুলো একেবারে গায় গায় লাগানো, মাঝে কোনো ফাঁক নেই।

2. প্রত্যেকটা rectangle একেকটা class-এর উপর দাঁড়িয়ে আছে। তাই class width-এর উপর নির্ভর করে rectangle-গুলোর প্রস্থের তারতম্য হতে পারে।

3. এখানে উচ্চতাগুলো হল relative frequency density.

এই ছবি দিয়ে statistical regularity ব্যবহার করে pdf আন্দাজ করার জন্য কী করব দ্যাখো। অনেক অনেক বার তীর ছুঁড়ব, এবং প্রত্যেকবার এইরকম histogram আঁকব। সেই সঙ্গে class-গুলোকে সূক্ষ্ম থেকে সূক্ষ্মতর করতে থাকব। এক সময়ে দেখা যাবে যে, ছবিগুলো ক্রমশঃ pdf-এর গিয়ে converge করছে।

আমাদের টার্গেটের উদাহরণ দিয়ে ব্যাপারটা বোঝা যাক। আমরা Fig 9-এর ভিত্তিতে Fig 10-এর histogram-টা পেয়েছিলাম। এবার আরো অনেক বার তীর ছোঁড়া হল এবং সেই সঙ্গে টার্গেটটাকেও আরো সূক্ষ্মভাবে ভাগ করলাম Fig 11-এর মত করে। এবার histogram আঁকলে পাব Fig 12-এর মত কিছু একটা। দেখতে একটা গোদা গোদা সিঁড়ির ধাপের মত দেখাচ্ছে। এবার আরও তীর ছুঁড়লাম এবং আরও সূক্ষ্মভাবে ভাগ করলাম Fig 13-এর মত, এবার histogram-এর ধাপগুলো আর তত গোদা গোদা নেই (Fig 14)। যদি আরো সূক্ষ্ম করি (Fig 15), তবে আর সিঁড়ি না বলে করাতের ফলা বললে বেশী ভালো হবে। সুতরাং যতই সূক্ষ্ম করছি ততই মসৃণতার দিকে এগিয়ে যাচ্ছে। বুঝতেই পারছ যে, শেষমেশ জিনিসটা Fig 16-এর মত একটা curve-এ গিয়ে converge করবে। এই curve-টাই হবে pdf. এই কথাটা আমরা এখানে প্রমাণ করব না। ■



Fig 13

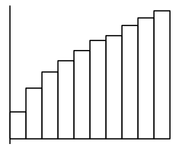


Fig 14

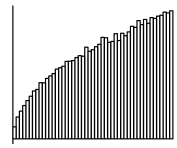


Fig 15

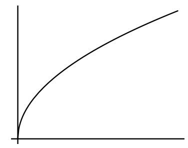


Fig 16

7.2.3 CDF আন্দাজ করার কায়দা

এবার আমরা দেখব random sample থেকে কীভাবে CDF আন্দাজ করা যায়। এক্ষুণি দেখলাম যে, PMF বোঝার জন্য লাগে barplot, আর pdf বোঝার জন্য histogram. তেমনি CDF বোঝার হাতিয়ার হল **empirical cumulative distribution function (ECDF),** বা সংক্ষেপে যাকে নীচের অংকটায় খালি empirical distribution বলেছে।

Example 13: Explain the term: empirical distribution. **(2014,2008)**

SOLUTION:

> **DEFINITION: Empirical distribution**
>
> If $X_1, ..., X_n$ are IID random variables then the empirical distribution is defined as the function $\widehat{F}_n : \mathbb{R} \to [0,1]$ where
>
> $$\forall x \in \mathbb{R} \quad \widehat{F}_n(x) = \frac{\text{number of times } X_i \leq x}{n}.$$

■

এই সংজ্ঞাটা পড়ে মাথামুণ্ডু কিছু বুঝতে পারা বেশ কঠিন। তাই একটা ছোটো উদাহরণ নেওয়া যাক। ধরো

$$X_1 = 3.4, \;\; X_2 = 1.1, \;\; X_3 = 4.1, \;\; X_4 = 2.0.$$

এখানে দেখতেই পাচ্ছ যে $n = 4$ নিয়েছি। আমরা $\widehat{F}_4(x)$ বার করব। এটা $x \in \mathbb{R}$-এর একটা function. প্রথমে যে কোনো একটা x নাও, যেমন ধরো $x = 3.5$. গুণে দ্যাখো কতগুলো X_i আছে যারা $\leq x$. উত্তর হল--তিনটে, X_1, X_2 আর X_4. এই 3 সংখ্যাটাকে ভাগ করব sample size দিয়ে, মানে $n = 4$ দিয়ে। সুতরাং

$$\widehat{F}_4(3.5) = \frac{3}{4}.$$

একইভাবে $\widehat{F}_4(2.0)$ হবে $\frac{2}{4}$. এইভাবে যেকোনো $x \in \mathbb{R}$-এর জন্যই $\widehat{F}_4(x)$ বার করা যাবে। তার জন্য প্রথমে Fig 17-এর মত একটা লাইনে সবগুলো X_i-কে বিন্দু দিয়ে দেখিয়ে নাও। যদি সবচেয়ে বাঁদিকের বিন্দুটারও বাঁদিকে কোনো x নাও, তবে $\widehat{F}_4(x) = 0$ হবে। প্রথম বিন্দু থেকে শুরু করে তার পর থেকে $\widehat{F}_4(x) = \frac{1}{4}$ হবে। সেটা চলতে থাকবে তার পরবর্তী বিন্দু অব্ধি। সেইখান থেকে শুরু করে $\widehat{F}_4(x) = \frac{2}{4}$ হবে। এইভাবে প্রতিটা বিন্দুতে এসে $\widehat{F}_4(x)$ লাফিয়ে লাফিয়ে বাড়বে। সব মিলিয়ে আমরা পাব একটা সিঁড়ির ধাপের মত গ্রাফ, যেমনটা দেখিয়েছি Fig 18-এ। ভালো করে গ্রাফটা বুঝে নাও। প্রত্যেকটা লাফের পরিমাণ হল $\frac{1}{4}$ করে, কারণ এখানে $n = 4$.

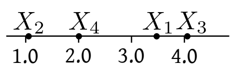


Fig 17

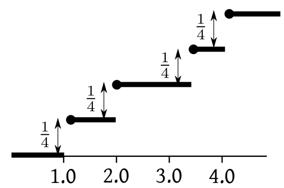


Fig 18

Empirical distribution-এর প্রধান গুরুত্ব এই যে, এটা ক্রমশঃ গিয়ে CDF-এ converge করে। কেন করে সেটা বোঝা খুবই সহজ। যেকোনো একটা সংখ্যা $a \in \mathbb{R}$ নাও। তাহলে

$$\widehat{F}_n(a) = \frac{\text{কত বার } X_i \leq a \text{ হয়েছে}}{n}$$

মানে হল $\{X \leq a\}$-এর relative frequency, যেটা converge করে $P(X \leq a)$-তে। আর এই $P(X \leq a)$-ই তো $F(a)$. সুতরাং $\forall a \in \mathbb{R} \quad \widehat{F}_n(a) \to F(a)$.

Example 14: Distinguish between distribution of a population and distribution of a sample.

"Distribution of the sample is the statistical image of the distribution of the population."–explain explicitly.[4] **(2004,2011)**

SOLUTION:

Let $X_1, ..., X_n$ be IID with some distribution with CDF F. Then--

Distribution of population...	Distribution of sample...
..is the distribution with CDF $F(x)$	...is the distribution with CDF $\widehat{F}_n(x) = \frac{1}{n}\|\{i \ : \ X_i \leq x\}\|$.
..may be any distribution	...must be a discrete distribution with finite sample space.
...is not random	...is random.

For example, if a fair coin is tossed independently 10 times to produce 3 heads and 7 tails, then

population distribution:

x	$P(X=x)$
head	$\frac{1}{2}$
tail	$\frac{1}{2}$

sample distribution:

x	$P(X=x)$
head	$\frac{3}{10}$
tail	$\frac{7}{10}$

Second part:

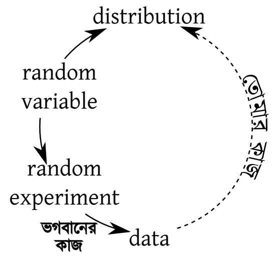


Fig 19

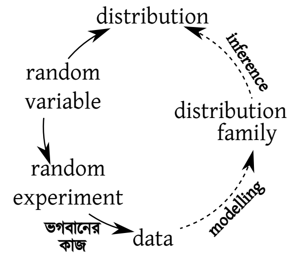


Fig 20

> It can be shown that the sample distribution converges to the population distribution as sample size $n \to \infty$. Thus, if we have a large sample, then the sample distribution gives a good approximation to the population distribution, just as a photographic image gives a good approximation to the object being photographed. For example, for a discrete population, the barplot of a large sample is close to the PMF of the population, For a population with pdf, the sample histogram converges to the population pdf.

■

7.3 দুই ধাপে করা

এই যে কাজটা, মানে data থেকে distribution অব্ধি পৌঁছোনো (Fig 19), সেটা দুই ধাপে করা হয়। কেন বলি। দেখা যায় যে, বহু বাস্তবক্ষেত্রে barchart বা histogram-এর কয়েকটা আদল বারবার করে আসে। হয়তো দুটো সম্পূর্ণ আলাদা random experiment, আলাদা random variable, অথচ আদলটা আশ্চর্যরকমের এক! এমনটা কেন হয় বলা শক্ত, কিন্তু হয় যে সেটা অনস্বীকার্য। একইরকম আদলের ভিত্তিতে distribution-গুলোকে কয়েকটা family-তে ভাগ করা হয়। একই family-র বিভিন্ন সদস্যর আদল একইরকম হয়, আবার আলাদা আলাদা family-র আদলও আলাদা আলাদা হয়। এইরকম বিভিন্ন পরিচিত family-দের একটা তালিকা আছে। তাদের আদল statistician-দের জানা থাকে।
এইবার ধাপদুটো কী বলি। প্রথম ধাপে barplot বা histogram-এর আদলটুকু খালি দেখে নেওয়া হয়, সেই আদলটাকে মনে মনে বিভিন্ন পরিচিত family-র চেহারার সঙ্গে মিলিয়ে সঠিক family-টা সনাক্ত করা হয়। এইবার দ্বিতীয় ধাপে আমরা চেষ্টা করি সেই family-র সঠিক সদস্যে পৌঁছোতে। প্রথম ধাপটাকে সাধারণতঃ বলে **modelling** আর দ্বিতীয় ধাপটাকে বলে **inference.** ব্যাপারটা অনেকটা Fig 20-এর মত।
এখানে একটা সমস্যা হতে পারে। হয়তো দেখলে যে, তোমার barchart বা histogram-এর আদলটা কোনো family-র সঙ্গেই খাপ খাচ্ছে না! তবে? সেক্ষেত্রে কাজটা একটু বেশী কঠিন। এই ধরণের সমস্যার মোকাবিলা করার জন্য statistics-এর একটা আস্ত শাখাই আছে, তার নাম nonparametric statistics. কিন্তু যদি সৌভাগ্যক্রমে আদলটা কোনো একটা family-র সঙ্গে মিলে যায়, তবে কাজটা অনেক সহজ। Statistics-এর এই শাখাকে বলে parametric statistics. আমরা এ বইতে খালি এই শাখা নিয়েই আলোচনা করব।

7.4 Family of distributions

একইরকম দেখতে distribution-দের নিয়ে একেকটা family তৈরী। সুতরাং একই family-র বিভিন্ন distribution-এর PMF বা pdf-দের একই ফর্মুলা দিয়ে লেখা হয় সেই ফর্মুলার মধ্যে কিছু **parameter** থাকে যা দিয়ে বিভিন্ন সদস্যদের আলাদা করা যায়। উদাহরণ দেখা যাক, তবে স্পষ্ট হবে।

Example 15: Fig 21-এ তিনটে density এঁকেছি। এদের সবারই আদল একইরকম, 0-র কাছে সবচেয়ে উঁচু, তারপর

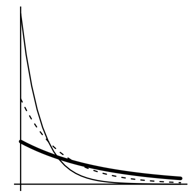


Fig 21

থেকে নেমে আসছে। যতই নামছে ততই বেশী আস্তে নামছে। এই আদলটাকে ধরা যায় এই ফর্মুলা দিয়ে

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

এখানে $\lambda > 0$ হল আমাদের parameter, যেটা দিয়ে এই family-র বিভিন্ন সদস্যদের আলাদা করে সনাক্ত করা যাবে। যেমন ধরো $\lambda = 1$ নিলে পাবে Fig 21-এর ড্যাশ্ ড্যাশ্ দাগ দিয়ে আঁকা density-টা। যদি $\lambda = 2$ নাও, তবে পাবে সরু দাগটা, আর $\lambda = 0.5$ নিলে পাবে মোটাটা। আমরা এখানে বলব

$$\{f_\lambda \;:\; \lambda > 0\}$$

হল একটা family of distributions. এই পরিবারের প্রতিটা সদস্যকে বলে এককেটা exponential distribution. লেখার সময়ে লিখি Expo(λ). সুতরাং যদি কখনো লেখা দ্যাখো যে

$X_1, ..., X_n$ are IID Expo(λ) for some $\lambda > 0$

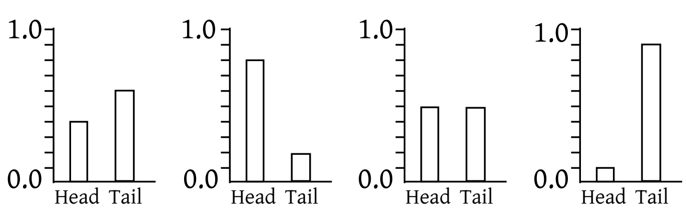
তবে বুঝবে যে X_i-দের density এই family-র অন্তর্ভুক্ত, তবে ঠিক কোন সদস্য সেটা এখনও সনাক্ত করা হয় নি, তাই খালি $\lambda > 0$ লিখে ছেড়ে রাখা হয়েছে। ■

এই family-টা ছিল continuous random variable-দের জন্য। এবার একটা discrete উদাহরণ দেখি।

Example 16: একটা কয়েন যদি বারবার করে টস করো, তবে কিছু head আর tail পাবে। এদের দিয়ে barchart আঁকলে সেটা হবে Fig 22-এর ছবিগুলোর মত কিছু একটা। এরা প্রত্যেকেই এককেটা probability mass function. প্রতিক্ষেত্রেই ঠিক দুটো করে value সম্ভব, head আর tail. তাই sample space-এর সাইজ হল 2. পার্থক্য হল খালি probability-গুলো নিয়ে। সুতরাং এই জাতীয় যাবতীয় PMF-দের চেহারা হল--

$$p_\theta(x) = \begin{cases} \theta & \text{if } x = \text{head} \\ 1-\theta & \text{if } x = \text{tail} \\ 0 & \text{otherwise} \end{cases}$$

Fig 22



যেখানে $\theta \in [0, 1]$. এই family-টার নাম হল Bernoulli family, এর সদস্যদেরকে Bernoulli(θ) নামে লেখা হয়। ■

Example 17: Statistics-এর সবচেয়ে বহুল ব্যবহৃত যে family of distributions, তার নাম normal বা Gaussian. এই family-র প্রত্যেকের PDF এই আদলের--

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

এখানে দুটো parameter আছে, $\mu \in \mathbb{R}$ আর $\sigma > 0$. চাইলে আমরা দুটো parameter-কে এক সঙ্গে (μ, σ) হিসেবেও লিখতে পারি। সেক্ষেত্রে parameter space হবে $\mathbb{R} \times (0, \infty)$. এই family-টার প্রতিটা সদস্যকে সাধারণতঃ $N(\mu, \sigma)$ বা Normal(μ, σ) হিসেবে লেখা হয়। কয়েকটা উদাহরণ রয়েছে Fig 23-এ। ■

তোমাদের মধ্যে যারা কম্পিউটার নাড়াচাড়া করতে ভালোবাসো তাদের পক্ষে হয়তো এইভাবে ভাবলে সুবিধা হবে--মনে করো PDF বা PMF-এর গ্রাফটা আসলে কম্পিউটারের পর্দায় একটা অ্যানিমেশন। প্রতিটা parameter-এর জন্য একটা করে স্লাইডার আছে যেটাকে তুমি মাউস দিয়ে সরাতে পারো (Fig 24)। যেমন যেমন সরাবে, parameter-এর value-ও তেমনি পরিবর্তিত হবে, আর গ্রাফটাও তাল রেখে বদলাবে। অবশ্য অ্যানিমেশন-ট্যানিমেশন কম্পিউটারের পর্দায় যত ভালো বোঝা যায়, বইয়ের পাতার নিথর ছবির দিকে তাকিয়ে সেটা কল্পনা করা কঠিন। তাই যাদের কম্পিউটার আছে (বা স্মার্ট ফোন) তাদের জন্য এরকম কিছু অ্যানিমেশন বানিয়ে রেখেছি এই পাতায় `http://www.isical.ac.in/~arnabc/probstat/`বাকিদের জন্য কল্পনাশক্তিই ভরসা।

নানা ধরণের family সম্ভব। তাদের কয়েকজনকে তো দেখলে, বাকিদের সঙ্গে যথাসময়ে পরিচয় হবে। আপাততঃ একটা নতুন ভাষা শিখে রাখি। Parameter-গুলো যা যা value নিতে পারে তাদের set-টাকে বলে **parameter space.** যেমন Expo(λ)-র বেলায় আমরা বলেছিলাম $\lambda > 0$ হতে হবে (নইলে PDF-এর ফর্মুলা বলে যেটা দেওয়া হয়েছিল সেটা আর PDF থাকবে না)। তার মানে এখানে parameter space হল $(0, \infty)$.

7.5 Modelling-এর মূল্য

<u>ঘটিরাম</u>: চার সের আলুর দাম যদি দশ আনা হয়, তবে আধ মণ পটলের দাম কত?
<u>পণ্ডিত</u>: ... আহা, আবার পটল এল কোত্থেকে?
<u>ঘটিরাম</u>: তা তো জানি না। বোধহয় পটলডাঙা থেকে।
--ঝালাপালা (সুকুমার রায়)

Modelling মানে হল distribution-এর family-টাকে সনাক্ত করা। খামোখা আমরা সেই কাজটা করছি কেন? Family-টা জেনে লাভটা কি হবে? এক কথায় এর উত্তর হল--অল্প data থেকে বেশী জ্ঞান আহরণ করা যাবে। দুটো উদাহরণ দেখা যাক। প্রথমটায় খালি ধারণাটা বুঝব, কোনো অংক ছাড়া।

Example 18: একটা লোক জামা পড়েছে Fig 25-এর উপরের অংশে দেখানো জামাটার মত। মত। খালি এই তথ্যের

Fig 23

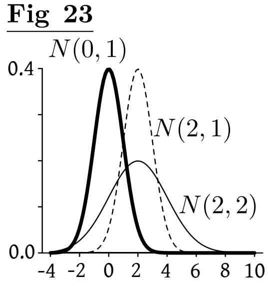
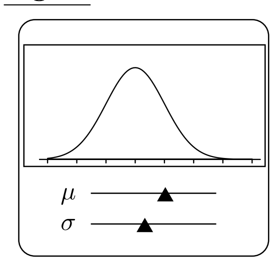


Fig 24



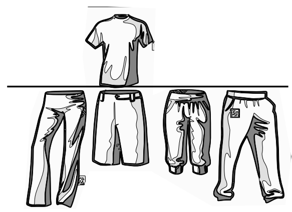


Fig 25

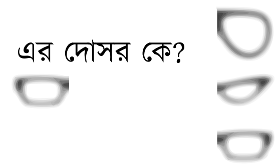


Fig 26

ভিত্তিতে বলো তো তার প্যান্টটা দেখতে নীচের অংশের ছবিগুলোর কোনটার মত? একটা লোকের চশমার অর্ধেকটা দেখতে Fig 26-এর বাঁদিকের ছবিটার মত। বল তো বাকি অর্ধেকটা ডানদিকের এর কোনটার মত?
SOLUTION: প্রথম ক্ষেত্রে জোর দিয়ে কিছু বলা মুস্কিল। কিন্তু দ্বিতীয় ক্ষেত্রে অবশ্যই শেষের ছবিটাই উত্তর হবে। এইটা বলা গেল কারণ চশমা জিনিসটার একটা গুরুত্বপূর্ণ বৈশিষ্ট্য হল যে, সেটা ডাইনে-বাঁয়ে symmetric হয়। এই আদলটুকু জানা ছিল বলেই খালি একটা অর্ধেক দেখেই বাকি অর্ধেকের বিষয়ে জ্ঞান আহরণ করা গেল। ■

এইবার এই একই ধারণার গাণিতিক রূপ দেখবে নীচের উদাহরণে।

Example 19: একটা random variable আছে X. বলা আছে যে $P(X<2)=\frac{1}{2}$. খালি এটুকু শুনেই বলতে পারো $P(X>1)$ কত হবে? এবার ধরো তোমার জানা আছে যে X-এর distribution-টা হল Expo(λ), যেখানে $\lambda>0$ তোমার অজানা। এবার উত্তর দিতে পারবে?
SOLUTION: প্রথম ক্ষেত্রে কিছুই বলা যায় না। কিন্তু দ্বিতীয় ক্ষেত্রে আমরা density-র চেহারাটা জানি এইরকম--

$$f_\lambda(x)=\begin{cases}\lambda e^{-\lambda x} & \text{if } x>0\\ 0 & \text{otherwise}\end{cases}$$

সুতরাং

$$P(X<2)=\int_{-\infty}^{2} f_\lambda(x)dx=\int_0^2 \lambda e^{-\lambda x}dx=1-e^{-2\lambda}=\frac{1}{2},\ \text{(দেওয়া ছিল)।}$$

অর্থাৎ $\lambda=\frac{1}{2}\log 2$. সুতরাং

$$P(X>1)=\int_1^\infty \lambda e^{-\lambda x}dx=e^{-\lambda}=\frac{1}{\sqrt{2}}.$$

লক্ষ কর যে, আমি কিন্তু তোমাকে X-এর distribution পুরো বলে দিই নি, খালি family-টুকু বলেছিলাম, কারণ λ-র value আমি বলি নি। কিন্তু তা থেকেই তুমি λ-র value-টা বার করে অংকটা করে দিতে পারলে। সুতরাং এমনিতে $P(X<2)$ আর $P(X>1)$-এর মধ্যে কোনো সম্পর্ক না থাকলেও family-টা জানলে λ ওদের মধ্যে একটা যোগসূত্র রচনা করল। $P(X<2)$ থেকে তুমি λ-য় পৌঁছলে, সেখান থেকে $P(X>1)$-এ। এইখানেই modelling-এর সুবিধা। ■

DAY 8
Estimation

Fig 27 আমরা আগেই দেখেছি। এখানে statistician-এর কাজটাকে দুভাবে ভাগ করে দেখানো হয়েছে--modelling আর inference. এবার আমরা একধরণের inference-এর কথা শিখব, যার নাম estimation. এটাকে ভগবানের সঙ্গে একটা খেলা বলে ভাবলে সুবিধা হবে--

> একটা random variable আছে (সুতরাং তার সঙ্গে একটা probability distribution এবং একটা random experiment-ও আছে)। তুমি probability distribution-টা জানো না। ভগবান random experiment-টা বার বার করে independently করেছেন এবং তোমাকে তার outcome-গুলো বলেছেন $X_1, ..., X_n$. তুমি barplot বা histogram এঁকে বুঝে ফেলেছো ভগবান কোন family নিয়ে কাজ করছেন (মানে modelling করা হয়ে গেছে)। তোমার কাজ এবার সেই family-র সঠিক সদস্যকে খুঁজে বার করা।

ব্যাপারটাকে অ্যানিমেশন দিয়ে ভাবলে সুবিধা হবে। একটু আগেই আমরা একটা family-র বিভিন্ন সদস্যকে অ্যানিমেশন দিয়ে ভাবছিলাম। সেখানে পর্দার উপর ছিল খালি p.d.f বা PMF-এর গ্রাফটা, আর প্রত্যেকটা parameter-এর জন্য একটা করে স্লাইডার। এবারও সেই সবই থাকবে, খালি বাড়তির মধ্যে পর্দার উপর histogram বা barplot-টাও আঁকা থাকবে। Fig 28 দ্যাখো। তোমার কাজটা হবে স্লাইডারগুলোকে এমনভাবে সরানো যাতে গ্রাফটা histogram-এর বা barplot-এর মাথাগুলো আলতো করে ছুঁয়ে যায়। যেমন ধরো Fig 28-এ মোটেই সেটা হয়নি। গ্রাফটাকে বাঁদিকে আরও সরানো দরকার। তার জন্য স্লাইডারগুলো নিয়ে খানিকক্ষণ নাড়াচাড়া করে দেখতে হবে। Fig 29-এ ছবিটা বেশ ঠিকঠাক মনে হচ্ছে। এরকম অবস্থায় স্লাইডারের থেকে যে value-গুলো পাবে তাদেরকেই বলব parameter-গুলোর **estimate.** কোনো parameter-র estimate বার করার প্রক্রিয়াটাকে বলে **estimation.** যদি দ্যাখো যে হাজার চেষ্টা করেও কাজটা করা যাচ্ছে না, তবে জানবে যে তোমার family নির্বাচনে ভুল হয়েছিল।
এই অ্যানিমেশনের কায়দাটা কল্পনা করার পক্ষে ভালো, কিন্তু এর মধ্যে একটা চোখের আন্দাজের প্রশ্ন থেকেই যায়। আমরা এবার একটা কায়দা শিখব যা দিয়ে একই কাজ অংক কষে করা যায়। সেই কায়দাটার মধ্যে তোমার data (মানে $X_1, ..., X_n$), এবং তোমার নির্বাচিত family-টা ঢুকিয়ে দিলেই একেবারে অংক কষে বেরিয়ে আসবে স্লাইডারগুলো কোনটা কোথায় রাখলে সবচেয়ে ভালো হবে। কায়দাটার নাম হল maximum likelihood estimation (MLE). এইটাই estimation-এর একমাত্র পদ্ধতি নয়, কিন্তু সবচেয়ে বহুলব্যবহৃত পদ্ধতি।

8.1 Maximum likelihood estimation (MLE)

ব্যাপারটা প্রথমে একটা গল্প দিয়ে বোঝা যাক।
হরবিলাসবাবু খুন হয়েছেন। সন্দেহভাজন তিনজন--

- এক, তাঁর বাড়ির চাকর, যে পাঁচদিন আগে কাজে বহাল হয়েছে,
- দুই, হরবিলাসবাবুর ভাইপো, যে অন্য পাড়ায় থাকে।
- আর তিন হল হরবিলাসবাবুর প্রতিবেশী গোপেন বিশ্বাস।

তদন্তকারী গোয়েন্দার হাতে তথ্য আছে এইরকম--

> খুনটা হয় সকাল দশটা নাগাদ। হরবিলাসবাবু একা মানুষ। সকাল আটটার সময়ে মর্নিং ওয়াক সেরে ফেরেন, সেদিনও ফিরেছিলেন। পাড়ার একাধিক লোক সাক্ষী আছে সে বিষয়ে। তারপর ঘন্টা দুই ধরে মনোযোগ দিয়ে বাংলা এবং ইংরাজী খবরকাগজ পড়েন। খুনটা হয়েছে সেই সময়েই। হরবিলাসবাবুর বাড়ির সামনে লোকচলাচল কম নয়, বিশেষতঃ একটা ক্লাব আছে, যেখানে ছেলে ছোকরারা হর্দম আড্ডা দিচ্ছে। তারা কেউ হরবিলাসবাবুর বাড়িতে সাড়ে আটটার পর আর কাউকে ঢুকতে বা বেরোতে দেখে নি। গোপেন বিশ্বাসের বাড়ি অবশ্য হরবিলাসবাবুর

Fig 27

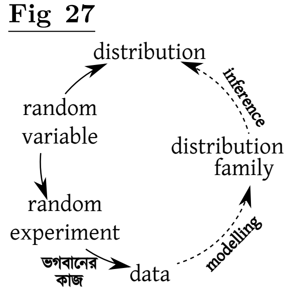


Fig 28

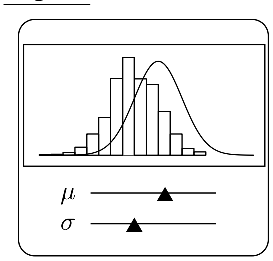
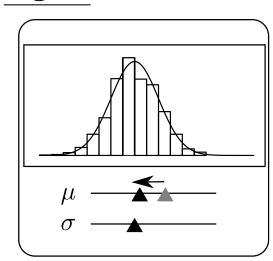


Fig 29



> সঙ্গে লাগোয়া, দেওয়ালের মধ্যে একটা ছোটো দরজা মতন আছে। সেখান দিয়ে গোপেনবাবু যদি ঢুকে থাকেন, তবে সেটা ক্লাবের ছেলেদের নজরে পড়ার কথা নয়। কিন্তু গোপেনবাবু সেদিন সকাল নয়টাতে বর্ধমান ছিলেন এ কথার সাক্ষী আছে।

এ হেন অবস্থায় সন্দেহের তীরটা কার দিকে পড়বে? উত্তর হল--চাকরটার দিকে। না, এইটুকু তথ্যের ভিত্তিতেই জোর দিয়ে বলা যাচ্ছে না যে, চাকরটাই খুনী। কিন্তু সন্দেহটা সবচেয়ে বেশী তার উপরেই আসতে বাধ্য, কারণ--

1. গোপেনবাবু যদি খুনী হন, তবে তাঁকে মাত্র এক ঘন্টার মধ্যে বর্ধমান থেকে কলকাতায় পৌঁছে খুনটা করতে হয়েছে। তার জন্য রাজ্যের পরিবহন ব্যবস্থা যতটা উন্নত হওয়া প্রয়োজন ততটা মোটেই নয়।

2. হরবিলাসবাবুর ভাইপো অবশ্য বেপাড়ায় থাকলেও দূরত্ব খুব বেশী নয়। ভাইপোর সঙ্গে সদ্ভাবও বিরাট কিছু ছিল না। কিন্তু একটা মানুষ প্রকাশ্য দিবালোকে হরবিলাসবাবুর বাড়িতে ঢুকেছে, অথচ ক্লাবের ছেলেরা কেউ তা দেখেছে বলে মনে করতে পারছে না, সেটা একেবারে অসম্ভব না হলেও সম্ভাবনা বেশ কম।

3. কিন্তু চাকরটার বেলায় এরকম কোনোই সমস্যা নেই।

সুতরাং সন্দেহটা চাকরটার উপরেই সবচেয়ে বেশী।
এই যুক্তিটা নিতান্তই সহজ বুদ্ধি। এতই সহজ যে, এ দিয়ে গোয়েন্দা গল্প জমবে না। কিন্তু বাস্তব জগতে হোম্‌স্ বা ফেলুদাসুলভ চাতুরীর চাইতে এই যুক্তিটাই অধিকাংশ অপরাধী ধরতে কাজে আসে বেশী। এই যুক্তিটাই হল maximum likelihood estimation-এর একটা প্রয়োগ। এখানে প্রকৃত খুনী যেন একটা parameter (θ). প্রকৃত খুনীর পরিচয় জানি না, খালি জানি যে সে তিনজনের মধ্যে একজন। এই সন্দেহভাজনদের set-টা হল parameter space. একে Θ নাম দেব। এখানে Θ হল বড় হাতের θ.

$$\Theta = \{\text{চাকর, ভাইপো, গোপেনবাবু}\}.$$

হাতে যা যা তথ্য ছিল (ক্লাবের ছেলেদের বয়ান, গোপেনবাবুর বর্ধমানে থাকা ইত্যাদি) সেগুলো মিলিয়ে হল data. আমরা এবার মনে মনে তিন সন্দেহভাজনের প্রত্যেককে খুনীর জায়গায় দাঁড় করিয়ে দেখলাম যে, যা যা ঘটেছে সেগুলো ঘটার সম্ভাবনা কোন ক্ষেত্রে কতটা। গোপেনবাবুর ক্ষেত্রে সম্ভাবনা প্রায় ছিল না বললেই হয়। ভাইপোর ক্ষেত্রেও সম্ভাবনা বেশ কম। কিন্তু চাকরের ক্ষেত্রে অনেকটাই বেশী। আমরা যেটাকে সম্ভাবনা বলছি, সেটাকেই statistics-এ বলে likelihood[3]। এবং সেই জিনিসটা কোথায় maximum সেটা দেখা হচ্ছে, তাই maximum likelihood estimation নাম। আমরা likelihood-কে সাধারণতঃ $L(\theta)$ লিখে থাকি। যেমন এখানে $L(\text{গোপেনবাবু})$ খুব কম, $L(\text{ভাইপো})$ তার চেয়ে বেশী, কিন্তু $L(\text{চাকর})$ সবচেয়ে বেশী। যে θ-র জন্য $L(\theta)$ সবচেয়ে বেশী সেইটাকেই বলে θ-র maximum likelihood estimate বা MLE. একে আমরা লিখি $\hat{\theta}$. যেমন আমাদের ক্ষেত্রে $\hat{\theta}=$চাকর।
এবার এই গল্পটার একটা গাণিতিক সংস্করণ দেখা যাক।

Example 20: ধরো একটা কয়েনকে 5 বার টস করলাম (অবশ্যই independently)। তার ফলে পেলাম H,H,T,H,T.

এ থেকে তোমায় estimate করতে হবে এই কয়েনটার জন্য $P(\text{head})$ কত। এখানে MLE কীভাবে লাগাবে?
SOLUTION: এখানে আমাদের data হল $X_1, ..., X_5$ যেখানে X_i হল i-th টসের outcome, যেমন আমাদের ক্ষেত্রে X_1 হয়েছে H, তারপর X_2-ও H, তারপর X_3 হয়েছে T, ইত্যাদি। যেহেতু একই কয়েন বার বার করে independently টস করা হয়েছে, তাই $X_1, ..., X_5$ হবে IID এবং তাদের PMF হবে

$$p_\theta(x) = \begin{cases} \theta & \text{if } x = \text{head} \\ 1-\theta & \text{if } x = \text{tail} \\ 0 & \text{otherwise} \end{cases}$$

যেখানে $\theta = P(\text{head})$ হল আমাদের parameter. এখানে parameter space হল $[0,1]$, কারণ যে কোনো probability-ই থাকে $[0,1]$-এর মধ্যে। এবার আমরা এই parameter space-এর প্রতিটা element-এর জন্য data-টার likelihood বের করব। সেটা হবে

$$L(\theta) = \theta \times \theta \times (1-\theta) \times \theta \times (1-\theta) = \theta^3(1-\theta)^2.$$

[3]Likelihood আর probability দুটোই সম্ভাবনা বোঝায়। ওদের মধ্যে একটা সূক্ষ্ম পার্থক্য আছে। সে কথায় যথাসময়ে আসব।

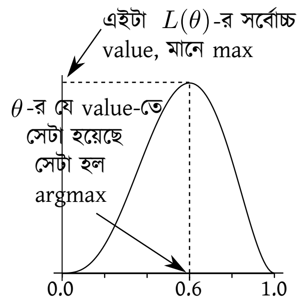


Fig 30

প্রতিটা head-এর জন্য একটা করে θ, আর প্রতিটা tail-এর জন্য $(1-\theta)$. সবগুলো গুণ হয়ে গেছে, কারণ কয়েন টসগুলো independent. এবার আমরা দেখব θ-র কোন value-র জন্য $L(\theta)$-টা maximum হয়। গ্রাফে আঁকলে $L(\theta)$-র ছবি হবে Fig 30-এর মত। যে θ-য় এটা maximum হয়েছে সেটাকে $\widehat{\theta}$ নাম দিয়েছি। এটাই হল θ-র MLE.
বার করার জন্য $L(\theta)$-কে differentiate করে দাও--

$$L'(\theta) = 3\theta^2(1-\theta)^2 - 2\theta^3(1-\theta).$$

আমরা $L(\widehat{\theta}) = 0$ বসিয়ে সমাধান করি--

$$3\widehat{\theta}^2(1-\widehat{\theta})^2 - 2\widehat{\theta}^3(1-\widehat{\theta}) = 0.$$

এ থেকে $\widehat{\theta}$-র তিনটে value পাচ্ছি, $0, 1$ এবং $\frac{3}{5}$. এর মধ্যে $L(0) = L(1) = 0$, কিন্তু $L\left(\frac{3}{5}\right) > 0$. সুতরাং 0 আর 1 এখানেই হিসেব থেকে বাদ গেল। এবার পড়ে রইল খালি $\frac{3}{5}$. এখানে যদি $L''\left(\frac{3}{5}\right)$ বার করো, দেখবে < 0 আসছে। সুতরাং MLE হল $\widehat{\theta} = \frac{3}{5}$. যেহেতু θ-র এই বিশেষ value-তে $L(\theta)$-টা maximum হয়, তাই আমরা একে বলি argmax–

$$\text{argmax}_{\theta\in[0,1]} L(\theta) = \tfrac{3}{5}.$$

Fig 30 দেখে নাও। Argmax-এর সংজ্ঞাটা মনে রেখো--যদি $f(x)$ কোনো function হয়, তবে $\text{argmax} f(x)$ মানে x-এর সেই value যার জন্য $f(x)$-টা maximum হবে। ■

ভালো করে লক্ষ করো এখানে মূল উপাদানগুলো কী কী। পয়লা নম্বর হল data. আমাদের বইয়ের সব অংকেই এই data হবে $X_1, ..., X_n$ জাতীয় যেখানে X_i-রা সবাই IID. যেহেতু IID, তাই সবগুলো X_i-এরই একই distribution. এই distribution-টা কোন family-তে সেটাও জানা আছে। এই family-টাই হল দুই নম্বর উপাদান। বাস্তব প্রয়োগের ক্ষেত্রে family-টা সনাক্ত করা হয় barplot বা histogram এঁকে। কিন্তু তার জন্য চাই অনেক data এবং তাদের ভিত্তিতে barplot বা histogram আঁকার জন্য কম্পিউটার। যেহেতু পরীক্ষার স্বল্প পরিসরে এ দুয়ের কোনোটাই জোটানো সম্ভব নয়, তাই পরীক্ষা-মার্কা অংকগুলোতে family-টা হয় বলে দেওয়া থাকবে, আর নয়তো এমন একটা গল্প বলা থাকবে যা থেকে family-টা সহজেই বোঝা যায়। এই অংকটায় যেমন কয়েন টসের গল্পটা বলা ছিল, তাই আমরা বুঝে গেলাম যে family-টা কী হবে। এরকম আরও কিছু গল্পের কথা আমরা এই বইতে পরে শিখব।

Example 21: Explain the method of maximum likelihood estimate. [3] **(2006,2008)**

SOLUTION:

Method of MLE:

প্রথমে উপকরণগুলোর কথা বলে নিই--

> Let $X_1, ..., X_n$ be a random sample from a distribution belonging to a given family with PMF or PDF $f_\theta(x)$. Here θ is an unknown parameter, that belongs to some known parameter space Θ.

বেশ খটমট শোনাচ্ছে, না? আসলে ব্যাপারটা কিন্তু আগের অংকে যেটা করলাম সেটার মতই। সেখানে $X_1, ..., X_5$ ছিল একই কয়েন বারবার টস করে পাওয়া। তাই আমরা জানতাম যে ওদের প্রত্যেকেরই distribution একই, যেটা Bernoulli family-র অন্তর্গত। মানে $P(\text{head}) = \theta$ আর $P(\text{tail}) = 1 - \theta$. এই চেহারাটুকু জানা ছিল, কিন্তু θ-টা ছিল অজানা। এবার বলি MLE-র উদ্দেশ্য কী--

> Maximum likelihood estimation is a procedure to estimate θ based on $X_1, ..., X_n$.

এবার প্রক্রিয়াটা লিখতে হবে--

> We define the likelihood of the data as
>
> $$L(\theta) = \prod_{i=1}^{n} f_\theta(X_i) = f_\theta(X_1) \times \cdots \times f_\theta(X_n).$$

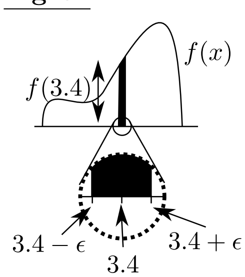
$\sum_1^n$ মানে যেমন পরপর যোগ করে যাওয়া, তেমনি $\prod_1^n$ মানে হল পরপর গুণ করে যাওয়া।
এইখানে একটা ব্যাপার একটু ভালো করে বোঝা দরকার। PMF-এর ব্যাপারটা আমরা আগের উদাহরণেই দেখেছি। এবার PDF-এর ব্যাপারটা দেখা যাক। ধরো $f_\theta(x)$ হল PDF, আর আমাদের data-তে $X_1 = 3.4$ আছে। তাহলে $P(X_1 = 3.4)$ কত হবে? উত্তর হল

$$P(X_1 = 3.4) = \int_{3.4}^{3.4} f_\theta(x)dx = 0.$$

সুতরাং এইভাবে এগোলে likelihood হয়ে যাচ্ছে 0. কিন্তু এখানে X_1 হল একটা continuous random variable. যখন আমরা বলছি $X_1 = 3.4$, তার আসল মানে কিন্তু হল $X_1 \in (3.4 - \epsilon, 3.4 + \epsilon)$, যেখানে ϵ আমাদের পরিমাপের সূক্ষ্মতা বোঝাচ্ছে। যেমন ধরো যদি আমরা সেন্টিমিটারে দৈর্ঘ্য মাপি, এবং আমাদের স্কেলে মিলিমিটারের ভাগ পর্যন্ত দেখানো থাকে, তবে 3.4 সেমি মানে হল এমন একটা দৈর্ঘ্য যার সবচেয়ে কাছের দাগটা হল 3 সেমি 4 মিমি। অর্থাৎ প্রকৃত দৈর্ঘ্য হবে 3.35 থেকে 3.45 সেমির মধ্যে কিছু একটা। এই কথাটা মাথায় রেখে probability বার করলে হবে

$$P\big(X_1 \in (3.4 - \epsilon, 3.4 + \epsilon)\big) = \int_{3.4_\epsilon}^{3.4+\epsilon} f_\theta(x)dx,$$

Fig 31



যেটা Fig 31-এর কালো জায়গাটার area-টা। যদি আমাদের যথেষ্ট সূক্ষ্মভাবে মেপে থাকি, তবে ϵ বেশ ছোটো হবে, এবং কালো জায়গাটা হবে খুব সরু। এতই সরু যে ওইটুকু জায়গার উপরে PDF-টাকে সর্বত্রই প্রায় $f(3.4)$-এর সমান বলেই মনে করা যায়। তাই ওই area-টা হবে মোটামুটি $f(3.4) \times 2\epsilon$.
একইরকম ব্যাপার হবে প্রতিটা X_i নিয়েই। সুতরাং মোট probability হবে মোটামুটি

$$(f_\theta(X_1) \times 2\epsilon) \times \cdots (f_\theta(X_n) \times 2\epsilon) = f_\theta(X_1) f_\theta(X_2) \cdots f_\theta(X_n) \times (2\epsilon)^n.$$

এখানে ϵ-এর মধ্যে কোনো θ নেই, ওটা কেবল আমাদের পরিমাপের সূক্ষ্মতার উপর নির্ভর করে। তাই maximise করার সময়ে আমরা খালি $f_\theta(X_1) \cdots f_\theta(X_n)$ নিয়েই কাজ করি, এবং এটাকেই likelihood function বলে থাকি। এখানেই likelihood আর probability-র পার্থক্য।

> Then the maximum likelihood estimate of θ is $\widehat{\theta} \equiv \widehat{\theta}(X_1, ..., X_n) \in \Theta$ such that
> $$\hat{\theta} = \text{argmax}_{\theta \in \Theta} L(\theta),$$
> if it exists.

অবশ্য এই maximisation-টা যে করা যাবেই এমন কোনো কথা নেই। বা করা গেলেও উত্তরটা unique নাও হতে পারে। সেই কথাটাও বলে দেওয়া ভালো।

> In general, this argmax may not exist or be unique.
>
> The MLE, $\widehat{\theta}$, (if it exists) may be considered as a value of the parameter that makes the observed data most likely.

একটা উদাহরণ দিয়ে দেওয়া সব সময়েই ভালো--

> For example, if a coin with unknown $P(\text{head}) = \theta \in [0, 1]$ is tossed twice independently, to produce the outcomes H, T, then the likelihood is
> $$L(\theta) = \theta(1 - \theta).$$
> This has maximum at $\widehat{\theta} = \frac{1}{2}$, which is the MLE of θ.

■

এবার আমরা MLE বার করার কিছু অংক করব। প্রতি ক্ষেত্রে আমাদের একটা random sample দেওয়া থাকবে $X_1, ..., X_n$, আর একটা family দেওয়া থাকবে। সেই family-র parameter-গুলোর MLE বার করাই হবে আমাদের কাজ। প্রথমে normal distribution নিয়ে কিছু উদাহরণ দেখি।

8.1.1 MLE with normal data

Example 22: Considering a sample of unit size from the population for a normal distribution (m, σ), find the maximum likelihood estimate of m assuming σ is known.[5] **(2005.8c)**
SOLUTION: আমরা random sample বলতে সাধারণতঃ $X_1, ..., X_n$ কল্পনা করি, যেখানে n হল sample size. এখানে sample of unit size বলেছে, মানে $n = 1$. সেটাকে আর X_1 না বলে X-ই বলা যাক--

> Let the data be $X \sim N(m, \sigma)$ where $m \in \mathbb{R}$ is unknown and $\sigma > 0$ is known.
> The likelihood function is
> $$L(m) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(X-m)^2}{2\sigma^2}\right).$$

এইটাই হল $N(m, \sigma)$-র density. এখানে parameter খালি m, যেহেতু σ জানাই আছে। তাই likelihood-টাকে লিখেছি m-এর function হিসেবে। এবার আমাদের কাজ হল এটাকে $m \in \mathbb{R}$-এর সাপেক্ষে maximise করা। এর জন্য আমরা একটা কৌশল করব, যাতে ব্যাপারটা অনেকটাই সহজ হয়ে যাবে। আমরা $L(m)$-কে maximise না করে $\log(L(m))$-কে maximise করব। যেহেতু log হল একটা strictly increasing function, তাই এতে $L(m)$-কেও maximise করা হয়ে যাবে। সাধারণতঃ $\log(L(m))$-এর চেহারা $L(m)$-এর থেকে সহজতর হয়।

> $\because \log x$ is a strictly increasing function of x,
> $\therefore$ maximising $L(m)$ is equivalent to maximising $\ell(m) = |log(L(m)) = -\frac{1}{2}\log(2\pi) - \log(\sigma) - \frac{1}{2s^2}(X-m)^2$.
> $\because \sigma^2 > 0$, $\therefore$this is same as minimising $(X-m)^2$ w.r.t. $m \in \mathbb{R}$.

এই জায়গাটা ভালো করে বুঝে নাও।

> Clearly, $\forall m \in \mathbb{R} \quad (X-m)^2 \geq 0$ and at $m = X$ we have $(X-m)^2 = 0$.
> So the minimum value of $(X-m)^2$is attained at $m = X$.
> Thus, the maximum value of $L(m)$ is attained at $m = X$.
> So the required MLE is $\widehat{m} = X$.

■

এবার তোমার হাত পাকানোর জন্য নীচের অংকটা দিলাম।

Exercise 4: Let $X_1, ..., X_n$ be IID $N(m, \sigma)$, where $m \in \mathbb{R}$ is unknown and $\sigma > 0$ is known. Then show that the MLE of m is $\widehat{m} = \bar{X} = \frac{1}{n}\sum_1^n X_i$.
HINT: এটাও ঠিক আগের অংকের মতই হবে। সেখানে যেমন $(X-m)^2$-কে minimise করতে হয়েছিল, এখানে তেমনি $\sum_1^n (X_i - m)^2$-কে $m \in \mathbb{R}$-এর সাপেক্ষে minimise করতে হবে। সে কাজটা differentiate করে করা যায়--

$$\frac{d}{dm}\sum_1^n (X_i - m)^2 = -2\sum_1^n (X_i - m) = 0,$$

মানে $\widehat{m} = \frac{1}{n}\sum_1^n X_i$. দ্বিতীয়বার differentiate করলেই দেখবে যে, second derivative-টা > 0 আসছে।
■

এই দুটো অংকে σ জানা ছিল, আমরা estimate করছিলাম m. এবার উল্টোটা করব, m জানা থাকবে, estimate করতে হবে σ বা σ^2.

Example 23: Find the maximum likelihood estimate of σ^2 of a normal(m, σ) population when m is known.[6] **(2008,2012)**
SOLUTION:

> Our random sample is
> $$X_1, ..., X_n \sim N(m, \sigma),$$
> where $m \in \mathbb{R}$ is known and $\sigma > 0$ is unknown.

এই যে লিখেছি $X_1, ..., X_n \sim N(m, \sigma)$, এর মানে হল $X_1, ..., X_n$-রা IID, এবন এদের প্রত্যেকেরই distribution হল $N(m\sigma)$. এই অর্থে "$\sim$" চিহ্নটা একটা প্রচলিত notation.
আমাদের estimate করতে হবে σ^2. তাই সেটাকে একটা নাম দিয়ে নিই--

> Let $\theta = \sigma^2$.
>
> The likelihood as a function of θ is
> $$\begin{aligned} L(\theta) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\tfrac{(X_i-m)^2}{2\theta}\right) \\ &= \left(\frac{1}{\sqrt{2\pi\theta}}\right)^n \exp\left(-\tfrac{1}{2\theta}\sum_1^n (X_i-m)^2\right). \end{aligned}$$

আবার সেই $\log$ নিয়ে সহজ করার কৌশলটা করব--

> Since $\log$ is a strictly increasing function, it is equivalent to maximising the log-likelihood $\ell(\theta) = \log L(\theta)$ w.r.t. θ.
>
> Now
> $$\ell(\theta) = -\frac{n}{2}[\log(2\pi) + \log\theta] - \tfrac{1}{2\theta}\sum_1^n (X_i-m)^2.$$
> $\therefore \ell'(\theta) = -\frac{n}{2\theta} + \frac{1}{\theta^2}\sum_1^n (X_i-m)^2$ Solving $\ell'(\widehat{\theta}) = 0$, we get
> $$\widehat{\theta} = \frac{1}{n}\sum_1^n (X_i-m)^2.$$
> The second derivative is
> $$\ell''(\theta) = \frac{n}{2\theta^2} - \frac{2}{\theta^3}\sum_1^n (X_i-m)^2.$$
> $\therefore \ell''(\widehat{\theta}) = \cdots < 0.$
>
> ওই ডট্ ডট্ অংশটুকু তুমি করে নিও।
>
> So $\widehat{\theta}$ <u>maximises</u> the log-likelihood function.

Thus the required MLE is

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_1^n (X_i - m)^2.$$

■

এইবার একটা অংক করব যেখানে একাধিক parameter আছে unknown.

Example 24: A random sample $(x_1, x_2, ..., x_n)$ is drawn from a normal(m, s) population. Find the maximum likelihood estimate of parameters m, s.[4] **(2004)**

Solution:

Our random sample is

$$X_1, ..., X_n \sim N(m, s),$$

where $m \in \mathbb{R}$ and $s > 0$ are both unknown.

The likelihood function is

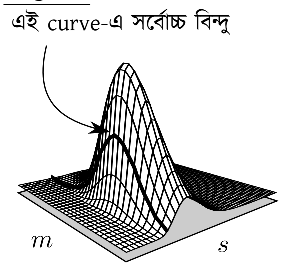
$$\begin{aligned} L(m,s) &= \prod_{i=1}^n \frac{1}{s\sqrt{2\pi}} \exp\left(-\tfrac{(X_i-m)^2}{2s^2}\right) \\ &= \left(\frac{1}{s\sqrt{2\pi}}\right)^n \exp\left(-\tfrac{1}{2s^2}\sum_1^n (X_i-m)^2\right). \end{aligned}$$

আবার সেই log নেওয়ার কৌশল--

Since log is a strictly increasing function, it is equivalent to maximising the log-likelihood

$$\ell(m,s) = \log L(m,s) = -\frac{n\log(2\pi)}{2} - n\log s - \tfrac{1}{2s^2}\sum_1^n (X_i-m)^2.$$

Fig 32



এখানে m আর s দুজনের সাপেক্ষে কী করে maximum বার করব বলি। একটা উপমা নাও--ধরো জানতে চাই ভারতের সেরা দৌড়বীর কে। সে কাজটা দু ধাপে করা যায়। প্রথমে প্রতিটা রাজ্যের মধ্যে একটা করে প্রতিযোগিতা করে বার করব সেই রাজ্যের সেরা দৌড়বীর কে। তারপর দ্বিতীয় ধাপে সেই সেরাদের মধ্যে একটা প্রতিযোগিতা করে ঠিক করব সেরাদের সেরা কে। সেই হবে পুরো ভারতের সেরা। এখানেও আমরা সেইভাবে এগোব। প্রথমে প্রতিটা s-এর জন্য আলাদা করে সেরা m বার করব। Fig 32 দ্যাখো। এখানে $L(m,s)$-এর গ্রাফ রয়েছে। দেখতে একটা তাঁবুর মত। এর চূড়োতেই পৌঁছতে চাই আমরা, মানে কোন (m,s)-এর জন্য $L(m,s)$ সবচেয়ে বেশী হয়। তার জন্য প্রথম ধাপে যে কোনো একটা s নির্দিষ্ট করে নিচ্ছি, অমনি আর পুরো তাঁবুটা নিয়ে মাথা ঘামাতে হচ্ছে না, খালি একটা curve নিয়ে মাথা ঘামালেই চলছে, যেমনটা মোটা করে দেখিয়েছি Fig 32-এ। সেই curve-এর সর্বোচ্চ বিন্দুটা বার করে দেখব সেটা কোন (m,s)-এর জন্য হচ্ছে।

For each fixed s, the maximum w.r.t. m is attained at $m = \overline{X}$.

এইটা আমরা আগেই করেছিলাম, যেখানে m-এর MLE বার করছিলাম, σ জানা আছে ধরে নিয়ে। যুক্তিটা সংক্ষেপে আরেকবার দিয়ে নিই--

[[Because:

$\frac{\partial \ell(m,s)}{\partial m} = \frac{1}{s^2}\sum_1^n (X_i - m)$.

Solving $\frac{\partial \ell(m,s)}{\partial m} = 0$, we get $\widehat{m} = \frac{1}{n}\sum X_i$.

Also $\frac{\partial^2 \ell(m,s)}{\partial m^2} = -\frac{n}{s^2} < 0$.

]]

এই হল প্রথম ধাপের শেষ। প্রতিটা রাজ্যের সেরা নির্বাচন হয়ে গেল। আমরা দেখলাম যে, যে কোনো s-এর জন্য সেরা (m,s) হচ্ছে $\left(\overline{X}, s\right)$. এবার এই সেরাদের মধ্যে প্রতিযোগিতা হবে।

Putting $m = \widehat{m}$, the log-likelihood becomes a function of s alone.

$$\ell(s) \equiv \ell(\widehat{m}, s) = -\frac{n\log(2\pi)}{2} - n\log s - \tfrac{1}{2s^2}\sum_1^n (X_i - \overline{X})^2.$$

$\therefore \ell'(s) = -\frac{n}{s} + \frac{1}{s^3}\sum_1^n (X_i - \overline{X})^2$.

Solving $\ell'(\widehat{s}) = 0$, we get

$$\widehat{s} = \sqrt{\frac{1}{n}\sum_1^n (X_i - \overline{X})^2}.$$

The second derivative is

$$\ell''(s) = \frac{n}{s^2} - \frac{3}{t^4}\sum_1^n (X_i - \overline{X})^2.$$

$\therefore \ell''(\widehat{s}) = \cdots < 0$.

ওই ডট্ ডট্ অংশটুকু তোমার করার জন্য রইল।

So $\widehat{s}$ maximises $\ell(s)$.

Thus the required MLE is

$$(\widehat{m}, \widehat{s}) = \left(\overline{X},\ \sqrt{\frac{1}{n}\sum_1^n (X_i - \overline{X})^2}\right).$$

■

DAY 9
MLE-র আরও উদাহরণ

গতকাল আমরা MLE-র বেশ কয়েকটা উদাহরণ দেখেছি। সেগুলো সবই ছিল normal distribution-দের family নিয়ে। আজকে অন্য কিছু family নিয়ে কাজ করব। প্রতিক্ষেত্রেই মূল ব্যাপারটা একই থাকবে--আমাদের data হবে $X_1, ..., X_n$. এদের distribution-এর family-টা দেওয়া থাকবে। এক বা একাধিক parameter থাকবে, যাদের MLE আমরা বার করব।

Example 25: Find an MLE for μ in Poisson distribution.[4] **(2013.5d)**

SOLUTION: ওরে বাবা, Poisson distribution-টা আবার কী জিনিস? ওটা আসলে একটা family-র নাম। Poisson-এর উচ্চারণ হল পোয়াসঁ। এর বিষয়ে পরে আমরা বিশদভাবে জানব। আপাততঃ এটুকু জানলেই হবে যে, এর প্রতিটা সদস্যই একেকটা discrete distribution, যাদের PMF-এর চেহারা হল

$$f_\mu(x) = \begin{cases} e^{-\mu}\frac{\mu^x}{x!} & \text{if } x = 0, 1, 2, ... \\ 0 & \text{otherwise} \end{cases}$$

এখানে $\mu > 0$ হল একটা parameter.

Let our data be $X_1, ..., X_n$ IID from Poisson(μ) distribution, where $\mu \in (0, \infty)$ is our unknown parameter.

The PMF is

$$f_\mu(x) = \begin{cases} e^{-\mu}\frac{\mu^x}{x!} & \text{if } x = 0, 1, 2, ... \\ 0 & \text{otherwise} \end{cases}$$

So the likelihood is

$$L(\mu) = \prod_1^n f(X_i) = e^{-n\mu}\frac{\mu^{\sum X_i}}{\prod X_i!}.$$

এখানে $\sum X_i$-টা কিন্তু μ-এর উপরতলায় power হিসেবে রয়েছে। আমাদের উদ্দেশ্য হল এটাকে maximise করা, তাই differentiate করে 0 করতে চাই। কিন্তু এটাকে differentiate করলে খুব ভদ্র কিছু আসবে না। তাই আগে log নিয়ে নিই--

The log-likelihood is

$$\ell(\mu) = \log(L(\mu)) = -\sum \log(X_i!) - n\mu + \left(\sum X_i\right)\log(\mu).$$

Since $\log x$ is a strictly increasing function, so it is enough to maximise

$\ell(\mu)$ w.r.t. μ.
Now

$$\ell'(\mu) = -n + \frac{\sum X_i}{\mu}.$$

Solving $\ell'(\widehat{\mu}) = 0$, we have

$$\widehat{\mu} = \frac{1}{n}\sum X_i,$$

if $\sum X_i > 0$.

এইখানে একটা সমস্যা দেখা দিচ্ছে। যদি $\sum X_i = 0$ হয়ে যায়? সেটা হতে পারে খালি একটাই ক্ষেত্রে, যদি সবগুলো X_i-ই শূন্য হয় (যেহেতু X_i-রা সবাই ≥ 0)। সেই কেসটাকে আলাদা করে দেখা দরকার--

If $\sum X_i = 0$, then $\forall i \quad X_i = 0$, and so $\ell(\mu) = -n\mu$, which has no maximum for $\mu \in (0, \infty)$. Hence MLE does not exist in this case.

যদি $\sum X_i > 0$ হয়, তবে অবশ্য সেই ঝামেলা নেই। তখন আমাদের second derivative test করতে হবে--

If $\sum X_i > 0$, then

$$\ell''(\mu) = -\frac{\sum X_i}{\mu^2} < 0.$$

So $\ell(\mu)$ is maximised at $\mu = \widehat{\mu} = \frac{1}{n}\sum_1^n X_i$, which is the required MLE.

■

Example 26: Estimate the parameter α of a continuous population having density function $(1+\alpha)x^\alpha$, $0 < x < 1$, by the method of maximum likelihood.[3] **(2006.4c)**
SOLUTION:

Let $X_1, ..., X_n$ be a random sample from the density

$$f(x) = \begin{cases} (1+\alpha)x^\alpha & \text{if } x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

এখানে parameter হল α. কিন্তু parameter space-টা বলা নেই। একটু চিন্তা করলেই বুঝবে যে,

For this to be a density, we need $\alpha > -1$.

কারণ যদি $\alpha < -1$ হয় তবে ওই গোড়ার $(1+\alpha)$-টা < 0 হয়ে যাবে, ফলে PDF-টাও negative হয়ে যাবে, যেটা অসম্ভব। আর যদি $\alpha = -1$ হয়, তবে PDF-টা সর্বত্রই 0 হয়ে যাবে, সুতরাং ওর মোট integral আর 1 থাকবে না!

Then the likelihood function is

$$\begin{aligned} L(\alpha) &= f(X_1)\cdots f(X_n) \\ &= \left[(1+\alpha)X_1^{\alpha}\right] \times \cdots \times \left[(1+\alpha)X_n^{\alpha}\right] \\ &= (1+\alpha)^n \left[X_1 X_2 \cdots X_n\right]^{\alpha}. \end{aligned}$$

Since $\log x$ is a strictly increasing function, maximising the likelihood is equivalent to maximising the log-likelihood,

$$\ell(\alpha) = \log L(\alpha) = n\log(1+\alpha) + \alpha\sum_1^n \log X_i.$$

Solving $\ell'(\widehat{\alpha}) = 0$,

$$\frac{n}{1+\widehat{\alpha}} + \sum_1^n \log X_i = 0$$

or, $\cdots$

মাঝের ধাপ কটা তুমি করে নিও

or, $$\widehat{\alpha} = -1 - \frac{n}{\sum_1^n \log X_i}.$$

এবার দেখে নেওয়া উচিত যে এই $\widehat{\alpha}$-টা parameter space-এর মধ্যে আছে কিনা, মানে $\widehat{\alpha} > -1$ হচ্ছে কিনা। তা যদি না হয়, তবে তো আর এই $\widehat{\alpha}$ নিয়ে কাজ চলবে না!

$\because X_i \in (0,1)$,
$\therefore \log X_i \in (-\infty, 0)$,
$\therefore \frac{1}{n}\sum_1^n \log X_i \in (-\infty, 0)$,
$\therefore \frac{n}{\sum_1^n \log X_i} \in (-\infty, 0)$,
$\therefore -\frac{n}{\sum_1^n \log X_i} \in (0, \infty)$,
$\therefore -1 - \frac{n}{\sum_1^n \log X_i} \in (-1, \infty)$,
$\therefore \widehat{\alpha} \in (-1, \infty)$, which is the parameter space.

যাক, বাবা, নিশ্চিন্ত, $\widehat{\alpha}$-টা parameter space-এর মধ্যেই আছে!

Thus we can take
$$\widehat{\alpha} = -1 - \frac{n}{\sum_1^n \log X_i}.$$

The second derivative at $\alpha = \widehat{\alpha}$ is
$$\ell''(\widehat{\alpha}) = -\frac{n}{(1+\widehat{\alpha})^2} < 0.$$

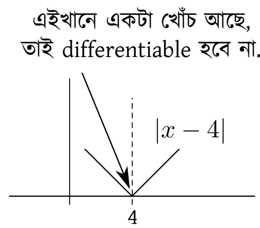


Fig 33

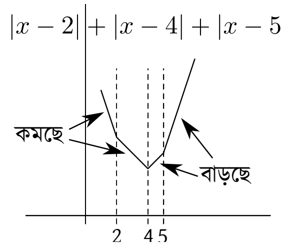


Fig 34

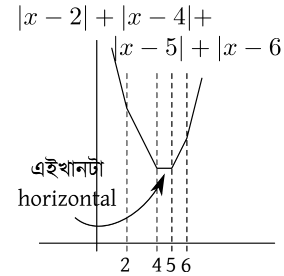


Fig 35

So we indeed have a maximum. Hence the required MLE is

$$\widehat{\alpha} = -1 - \frac{n}{\sum_1^n \log X_i}.$$

■

এবার একটা অংক দিচ্ছি তোমার চেষ্টা করার জন্য।

Exercise 5: A population is defined by the density function

$$f(x, \alpha) = \frac{x^{\ell-1} e^{-x/\alpha}}{\alpha^\ell \Gamma(\ell)}, \quad 0 < x < \infty,$$

ℓ being a known constant. Estimate the parameter α by the method of maximum likelihood.[4] **(2010)** ■

9.1 একটু কঠিন অংক

MLE ব্যাপারটা দেখতে গেলে খুব একটা কঠিন কিছু নয়, likelihood-টা লিখে ফেলে সেটাকে (বা তার log-কে) maximise করে ফেললেই হল! সমস্যা খালি হতে পারে ওই maximise করাটা নিয়েই। এবার সেরকম কিছু সমস্যার অংক দেখব।

9.1.1 যখন differentiate করা যায় না

এবার একটা উদাহরণ দেখব যেখানে likelihood function-টা (বা তার log) কেউই differentiable নয়। এর জন্য মনে করিয়ে দিই যে, $f(x) = |x-4|$ একটা differentiable function নয়। Fig 33 দ্যাখো। অবশ্য তা বলে এর minimum কোথায় হচ্ছে সেটা বার করতে কোনো অসুবিধা নেই, ছবি থেকে দেখাই যাচ্ছে যে, minimum-টা হচ্ছে $x = 4$-এ। কারণ $x < 4$ হলে $f(x)$-টা কমছে, আর $x > 4$ হলে বাড়ছে। যদি এরকম বেশ কয়েকটা জিনিসকে যোগ করা হয় তাহলেও একইরকম যুক্তি লাগানো যায়। ধরো $f(x) = |x-2| + |x-4| + |x-5|$. ছবিটা রয়েছে Fig 34-এ। লক্ষ কর যে এখানেও $f(x)$ প্রথমে নামছে, তারপর উঠছে, এবং minimum-টা আসছে $x = 4$-এ। এই ব্যাপারটা সবসময়েই হয়, যদি কিছু সংখ্যা থাকে $x_1, ..., x_n$ তবে $f(x) = \sum |x - x_i|$-এর minimum-টা হবে যখন x হবে x_i-গুলোর মধ্যে যেটা সবচেয়ে মাঝখানে আছে সেটা। যদি n একটা even (জোড়) সংখ্যা হয়, তবে সবচেয়ে মাঝখানে দুটো সংখ্যা থাকবে, তাদের মধ্যবর্তী যে কোনো জায়গাতেই $f(x)$-টা minimum হবে। যেমন ধরো যদি $n = 4$-টে সংখ্যা নিই-- $f(x) = |x-2| + |x-4| + |x-5| + |x-6|$, তবে যে কোনো $x \in [4,5]$-এর জন্যই minimum হবে। Fig 35 দ্যাখো।

এই তথ্যটা আমরা পরে প্রমাণ করব। আপাততঃ এটা ব্যবহার করব নীচের অংকে। একটা ভাষা শিখে রাখো-- $x_1, ..., x_n$ যদি কিছু সংখ্যা হয় তবে ওদের মধ্যে সবচেয়ে মাঝের সংখ্যাটাকে (বা n যদি even হয়, তবে সবচেয়ে মাঝের সংখ্যাদুটোর মধ্যবর্তী যেকোনো সংখ্যাকে) বলে $x_1, ..., x_n$-এর **median.**

Example 27: A random sample of size n has been drawn from a population of X with probability density function

$$f(x) = \text{const} \times e^{-\frac{|x-a|}{b}}, \quad b > 0, \quad -\infty < x < \infty.$$

Find the maximum likelihood estimate of population parameters a and b.[6] **(2009.4b)**

SOLUTION: এখানে দুটো parameter আছে a আর b. এরকম দুই parameter-ওয়ালা একটা অংক আগে করেছিলাম। কায়দাটা ছিল এইরকম--প্রথমে একটা parameter-কে স্থির রেখে অন্যটার সাপেক্ষে maximise করেছিলাম। যেটাকে উপমা দিয়েছিলাম প্রতিটা রাজ্যের মধ্যে সেরা দৌড়বীর নির্বাচনের সঙ্গে। তারপর সেই সেরাদের সেরাকে বার করেছিলাম। এখানেও একই কাজ করব।

> Let $X_1, ..., X_n$ be IID with the PDF
>
> $$f(x) = c \cdot e^{-\frac{|x-a|}{b}}, \quad -\infty < x < \infty.$$
>
> Here $a \in \mathbb{R}$ and $b \in (0, \infty)$ are the unknown parameters.

এখানে c কিন্তু একটা parameter নয়, কারণ $\int_{-\infty}^{\infty} f(x)dx = 1$ হওয়ায় c-কে a আর b দিয়ে প্রকাশ করা যাবে। সেই কাজটা আগে করে নিই--

> Now,
>
> $$\begin{aligned} \because \int_{-\infty}^{\infty} f(x)dx &= 1 \\ \therefore c \int_{-\infty}^{\infty} e^{-\frac{|x-a|}{b}} dx &= 1 \\ \therefore bc \int_{-\infty}^{\infty} e^{-|y|} dy &= 1 \quad \left[\text{putting } y = \tfrac{x-a}{b} \right] \\ \therefore 2bc \int_{0}^{\infty} e^{-y} dy &= 1 \quad [\because \text{ even function}] \\ \therefore 2bc &= 1. \end{aligned}$$
>
> Hence $c = \frac{1}{2b}$.

ব্যস্, c বেরিয়ে গেল। এবার likelihood-টা লিখে ফেলি--

> Then the likelihood function is
>
> $$\begin{aligned} L(a,b) &= f(X_1) \cdots f(X_n) \\ &= \left[\tfrac{1}{2b} e^{-\frac{|X_1-a|}{b}} \right] \times \cdots \times \left[\tfrac{1}{2b} e^{-\frac{|X_n-a|}{b}} \right] \\ &= (2b)^{-n} \exp\left[-\tfrac{1}{b} \sum_{1}^{n} |X_i - a| \right]. \end{aligned}$$
>
> Since $\log x$ is a strictly increasing function, maximising the likelihood w.r.t.

> a, b is equivalent to maximising the log-likelihood
>
> $$\ell(a,b) = \log L(a,b) = -n\log(2b) - \frac{1}{b}\sum_1^n |X_i - a|$$
>
> w.r.t. $a \in \mathbb{R}$ and $b > 0$.

এইবার দুই ধাপে এগোব। প্রথমে প্রতিটা $b > 0$-এর জন্য সবচেয়ে ভালো a বার করব।

> From standard result we know that $\sum_1^n |X_i - a|$ is the minimum when a is a median of the X_i. Thus, for each fixed $b > 0$, the log-likelihood function is maximum when $a = median(X_1, ..., X_n)$.

এইবার দ্বিতীয় ধাপে সেরাদের সেরা নির্বাচন হবে--

> Putting $a = \widehat{a}$ in the log-likelihood, we have a function of only b--
>
> $$\ell(b) = \ell(\widehat{a}, b) = -n\log(2b) - \frac{1}{b}\sum_1^n |X_i - \widehat{a}|.$$
>
> Thus
>
> $$\ell'(b) = -\frac{n}{b} + \frac{1}{b^2}\sum_1^n |X_i - \widehat{a}|.$$
>
> Solving $\ell'(\widehat{b}) = 0$,
>
> $$\widehat{b} = \frac{1}{n}\sum_1^n |X_i - \widehat{a}|.$$
>
> The second derivative at $b = \widehat{b}$ is
>
> $$\ell''(\widehat{b}) = \cdots < 0.$$

মাঝের ডট্ ডট্ অংশটুকু তুমি করে নিও।

> So indeed we have a maximum. Thus the required MLE is
>
> $$(\widehat{a}, \widehat{b}) = \left(median(X_1, ..., X_n), \tfrac{1}{n}\sum_1^n |X_i - median(X_1, ..., X_n)| \right).$$

■

9.1.2 Sample space-এর মধ্যেও parameter

এবার দুটো অংক দেখব যেখানে একটা নতুন প্যাঁচ আছে। Sample space-টা parameter-এর উপর নির্ভর করবে! কয়েন টসের বেলায় সে সমস্যা ছিল না, $P(\text{head})$ যাই হোক না কেন, sample space সর্বদাই একই থাকত $\{\text{head}, \text{tail}\}$. কিন্তু এবার নীচের অংকদুটোতে সেরকম হবে না।

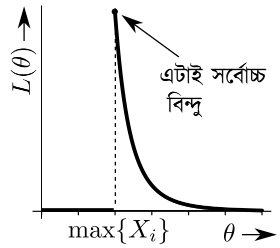


Fig 36

Example 28: Show that for a population of random variable X with probability density function $f(x) = 1/\theta$ for $0 \leq x \leq \theta$, the maximum likelihood estimator is the largest number of sample values.[5] **(2014.5d)**

SOLUTION: এখানে data কী সেটা বলে দিতে ভুলে গেছে। তাই আমরা ধরে নেব যে data হল একটা random sample, যাদের নাম দেব $X_1, ..., X_n$.
এখানে X_i-গুলো কী কী value নিতে পারে? উত্তর হল $[0, \theta]$-র মধ্যে যে কোনো value, অর্থাৎ sample space হল $[0, \theta]$, যেটা θ-র উপর নির্ভর করে।

> Let $X_1, ..., X_n$ be IID with PDF
>
> $$f(x) = \begin{cases} \frac{1}{\theta} & \text{if } x \in [0, \theta] \\ 0 & \text{otherwise} \end{cases}$$
>
> where $\theta \in (0, \infty)$ is unknown.
> The likelihood function is
>
> $$\begin{aligned} L(\theta) &= f(X_1) \cdots f(X_n) \\ &= \begin{cases} \left(\frac{1}{\theta}\right)^n & \text{if } \theta \geq \max\{X_i\} \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

এই ধাপটাতে অনেক ছাত্রছাত্রীই খাবি খায়। ওই $\max\{X_i\}$-টা এলো কোথা থেকে? লক্ষ কর যে, $L(\theta)$ এখানে $f(X_i)$-দের গুণফল। তাই একটাও $f(X_i)$ যদি 0 হয়ে যায়, তবে $L(\theta)$-ও 0 হয়ে যাবে। সুতরাং $L(\theta) > 0$ হওয়ার জন্য প্রতিটা $f(X_i) > 0$ হতে হবে, মানে সব i-এর জন্যই $X_i \leq \theta$ হতে হবে। এখন "θ সবগুলো X_i-এর থেকেই বড় বা সমান" বলা আর "θ সবচেয়ে বড় X_i-টার থেকে বড় বা সমান" বলা একই কথা। সেইভাবেই $\max\{X_i\}$-টা এসেছে।
$L(\theta)$-র গ্রাফটা আঁকা কঠিন নয় মোটেই (Fig 36)। এই গ্রাফটা আঁকাটা এই অংকে খুবই দরকারী। লক্ষ কর যে, likelihood-টা প্রথমে খানিকক্ষণ 0, তারপর এক লাফে খানিকটা উঠে ফের একটু একটু করে নেমে আসছে। নেমে আসছে, কারণ--

> Clearly, $\left(\frac{1}{\theta}\right)^n$ is a decreasing function of θ.

সুতরাং maximum-টা হচ্ছে ঠিক ওই লাফের জায়গাটাতেই।

> So the maximum is attained at $\widehat{\theta} = \max\{X_i\}$, which is the required MLE.

■

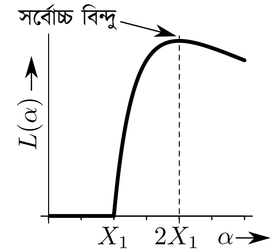


Fig 37

Example 29: Prove that the maximum likelihood estimate of the parameter α of the population having density function

$$f(x) = \frac{2(\alpha - x)}{\alpha^2}, \quad (0 < x < \alpha),$$

for a sample x_1 of unit size is $2x_1$.[5] **(2011)**

Let X_1 have PDF

$$f(x) = \begin{cases} \frac{2(\alpha-x)}{\alpha^2} & \text{if } x \in (0, \alpha) \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha \in (0, \infty)$ is unknown.

The likelihood function is

$$\begin{aligned} L(\alpha) &= f(X_1) \\ &= \begin{cases} \frac{2(\alpha-X_1)}{\alpha^2} & \text{if } \alpha > X_1 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

For $\alpha > X_1$ we have

$$L'(\alpha) = \frac{2\alpha^2 - 4\alpha(\alpha - X_1)}{\alpha^4} = \frac{-2(\alpha - 2X_1)}{\alpha^3}.$$

Clearly, for $\alpha > X_1$

$$L'(\alpha) \gtreqless 0 \text{ according as } \alpha \lesseqgtr 2X_1.$$

So $L(\alpha)$ attains its maximum value for $\alpha > X_1$ at $\widehat{\alpha} = 2X_1$.

Fig 37 দেখলে ব্যাপারটা স্পষ্ট হবে। লক্ষ কর যে, এখানে গ্রাফটায় একটা খোঁচ আছে। তাই পুরো $L(\alpha)$-টা কিন্তু differentiable নয়। সেই কারণেই আমরা খালি $\alpha > X_1$ নিয়ে কাজ করলাম (মানে খোঁচের ডানদিকটুকু নিয়ে)। এবার বাঁদিকের দিতে তাকাই। সেখানে তো সর্বত্রই $L(\alpha) = 0$.

Since $L(\widehat{\alpha}) > 0$, hence $\widehat{\alpha}$ is also the global maximiser.

So the MLE is $\widehat{\alpha} = 2X_1$, as required.

■

Answers

5. $\widehat{\alpha} = \frac{\sum_1^n X_i}{n\ell}$.

Chapter III
Conditional probability

DAY 10
Conditional probability

Probability আর statistics-এর কারবার হল random জিনিসপত্র নিয়ে। যদি বলি এরকম কোনো একটা জিনিসের উদাহরণ দিতে, তবে তার একটা চটজলদি উত্তর হল কয়েন টস। Head পড়বে নাকি tail পড়বে তার কোনো স্থিরতা নেই। আমি এক ভদ্রলোকের কথা জানি যিনি প্রথম জীবনে ম্যাজিসিয়ান ছিলেন, এবং যিনি একটা কয়েন টস করে ইচ্ছে মত head বা tail ফেলতে পারেন! না, এটা কোনো চোখে ধূলো দেওয়া কারসাজি নয়, বহু বছরের অভ্যাসের ফলে ওনার আঙুলের উপর এত নিয়ন্ত্রণ এসেছে যে, ঠিক কতটা জোরে কীভাবে পয়সাটাকে শূন্যে ছুঁড়ে দিলে head আসবে বা tail আসবে সেই হিসেবটা উনি জানেন। যে কয়েনটাকে টস করে তুমি probability বলছ $\frac{1}{2}, \frac{1}{2}$ করে, সেই কয়েনটাই এই ভদ্রলোকের হাতে পড়লে probability বেমালুম বদলে হয়ে যাবে $1, 0$ (বা $0, 1$)। সুতরাং লক্ষ কর যে, "একটা কয়েনের head পড়ার probability" আসলে কিন্তু কয়েনটার সম্পূর্ণ নিজস্ব বৈশিষ্ট্য নয়, যেমন নিজস্ব বৈশিষ্ট্য হল কয়েনটার ওজন বা তার সাইজ, বা অর্থমূল্য। Probability-টা আসলে প্রকাশ করছে ওর বিষয়ে আমাদের অজ্ঞতা কতটা। যে coin অজ্ঞজনের হাতে unbiased (অর্থাৎ $\frac{1}{2}, \frac{1}{2}$ করে probability) সেটাই ওস্তাদ লোকের হাতে ঘোরতর biased (মানে head আর tail-এর probability অসমান)। আমাদের জ্ঞানবৃদ্ধির সঙ্গে সঙ্গে কোনো ঘটনার probability কী করে পরিবর্তিত হয় সেটা ভালো করে বোঝার জন্য একটা হাতিয়ার আছে, তাকে বলে conditional probability, যেটা আমাদের এবারের আলোচ্য বিষয়।
একটা উদাহরণ দিয়ে শুরু করি।

Example 1: ইলেকট্রনিক ঘড়ি বা ক্যালকুলেটরে অনেক সময়ে সংখ্যা লেখা হয় Fig 1-এর মত করে। সাতখানা আলাদা

আলাদা আলো জুড়ে সংখ্যাগুলো তৈরী হয়, যাদেরকে Fig 2-এ দেখিয়েছি, $a, b, ..., h$ নাম দিয়ে। একটা ক্যাসিনোতে এরকম একটা বড় সংখ্যা বাক্সের মধ্যে রাখা আছে। একটা সুইচ টিপে দিলে তাতে একটা random সংখ্যা ফুটে ওঠে, সেটা 0 থেকে 9 যা খুশি হতে পারে। বাক্সের সামনে একটা দরজা বন্ধ থাকে, তাই সংখ্যাটা বাইরে থেকে দেখা যায় না। বাজি ধরা হয়, সংখ্যাটা জোড়[1] নাকি বিজোড়। যেহেতু দশটা সংখ্যার ভিতর ঠিক পাঁচটাই জোড়, তাই আর কোনো তথ্য হাতে না থাকলে

[1] মনে রেখো যে, 0 একটা জোড় সংখ্যা, কারণ 0-কে 2×0 আকারে লেখা যায়।

Fig 1

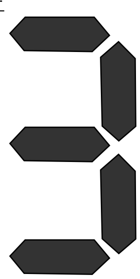
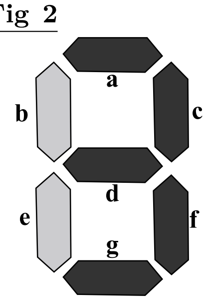


Fig 2



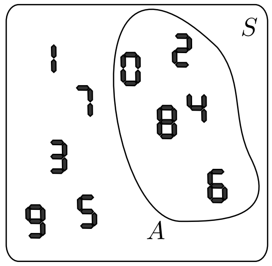


Fig 3

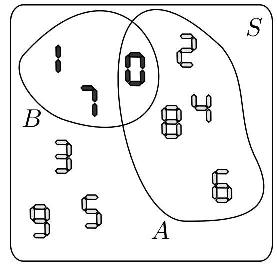


Fig 4

probability-টা হবে $\frac{5}{10} = \frac{1}{2}$. এবার ধরো দরজার মাঝামাঝি জায়গায় একটা ফুটো আছে। সেটা দিয়ে তুমি দেখে ফেলেছো যে, একদম মাঝের আলোটা জ্বলছে না। সুতরাং তুমি চট্‌ করে বুঝে গেলে যে সংখ্যাটা $2, 3, 4, 5, 6, 8$ বা 9 হতে পারে না, যেহেতু এদের বেলায় মাঝের আলোটা জ্বলে। অতএব পড়ে রইল $0, 1$ আর 7. এদের মধ্যে জোড় খালি 0. সুতরাং এই বাড়তি তথ্যটুকু জানার পর জোড়ের probability-টা হয়ে গেল $\frac{1}{3}$. ■

এই যুক্তিটাকে একটু তলিয়ে দেখি। গোড়াতে আমাদের sample space ছিল $S = \{0, 1, 2, ..., 9\}$, আর event-টা ছিল এদের মধ্যে যাবতীয় even সংখ্যাদের set, মানে $A = \{0, 2, 4, 6, 8\}$. এখানে Fig 3 দ্যাখো। যেই বাড়তি তথ্যটুকু ("মাঝের আলো জ্বলছে না") হাতে এল অমনি আমাদের sample space-টা সীমিত হয়ে গেল $B = \{0, 1, 7\}$-এ (Fig 4)। বাকিরা আমাদের চিন্তার বাইরে বেরিয়ে গেল (ছবিতে আবছা করে দেখিয়েছি)। ফলে A-ও সীমিত হয়ে দাঁড়ালো $A \cap B = \{0\}$. সুতরাং এখন প্রশ্নটা হল B-এর মধ্যে $A \cap B$ কতটা জায়গা দখল করে আছে। তাই উত্তর হবে $P(A \cap B)/P(B)$. এর একটা নাম আছে। একে বলে conditional probability of A given B. লেখার সময়ে লিখি $P(A|B)$.

Example 2: Define conditional probability $P(A|B)$.[1] **(2011,2009,2006,2014)**

SOLUTION:

> DEFINITION: **Conditional probability**
>
> For any two events A, B with $P(B) > 0$, the conditional probability of A given B is defined as
>
> $$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

এখানে দুটো কথা বলা দরকার--

- আমরা যখন $P(A|B)$ লিখছি, তখন Aআর B-এর মাঝখানে একটা খাড়া দাগ দিচ্ছি, ওটা হেলানো ভাগ চিহ্ন নয়, মানে $P(A/B)$ নয়। অনেককেই এই ভুলটা করতে দেখা যায়।
- সংজ্ঞায় একটা শর্ত রয়েছে যে, $P(B) > 0$. যেহেতু $P(B)$ দিয়ে ভাগ করা হচ্ছে, তাই $P(B) = 0$ হলে যে সমস্যা হবে সে তো বুঝতেই পারছ। কিন্তু ভাগের সমস্যা ছাড়াও আরেকটা সমস্যা হত সেটা বুঝতে কোনো অংক লাগে না। যদি $P(B) = 0$ হয়, তার মানে B ঘটা অসম্ভব। এদিকে তোমাকে বলা হয়েছে যে, B ঘটেছে! ঠিক যেমন কেউ যদি এসে বলে--"একটা ছক্কা দুইবার চালা হয়েছে, তাদের যোগ ফল হল 9.5, বল তো প্রথম দানে 6 পড়ার conditional probability কত?" তাহলে আমাদের পক্ষে হাঁ করে চেয়ে থাকা ছাড়া পথ নেই, কারণ যোগফলটা ভগ্নাংশ হয় কী করে?

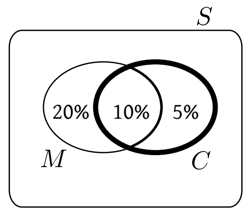


Fig 5

■

নীচের অংকটা সরাসরি এই সংজ্ঞার প্রয়োগ।

Example 3: In an examination 30% of the students failed in mathematics, 15% failed in chemistry and 10% failed in both mathematics and chemistry. A student is selected at random. If he has failed in chemistry, what is the probability that (i) he passed in mathematics, (ii) he has failed in both mathematics and chemistry?[5] **(2012.1e)**

SOLUTION: Fig 5 দ্যাখো। এখানে S হল যাবতীয় ছাত্রছাত্রীর set. এদের মধ্যে M হল যারা অংকে ফেল মেরেছে, আর C হল কেমিস্ট্রিতে ফেলটুসের দল। একটা random ছাত্র বা ছাত্রীকে নেওয়া হয়েছে। দেখা গেছে সে কেমিস্ট্রিতে ডাব্বা খেয়েছে। সুতরাং আমাদের নজর এখন C-এর দিকে, বাইরের সব কিছু আমাদের কাছে ঝাপসা।

> Let M and C be the events that the randomly selected student failed in mathematics and chemistry, respectively.
> Then it is given that $P(M) = 0.3, P(C) = 0.15$ and $P(M \cap C) = 0.1.$

আমরা প্রথমে দ্বিতীয় অংশটা করব। এখানে বার করতে হবে যে নির্বাচিত ছাত্র বা ছাত্রীটির অংকেও ধেড়ানোর probability, মানে $P(M|C)$. আমাদের নজর খালি C-এর দিকে, তার মধ্যে অংকে ফেল set-টা হল $M \cap C$, যার probability রয়েছে 10%. পুরো C-এর probability ছিল 15%. সুতরাং উত্তর হবে 15-এর মধ্যে 10 অর্থাৎ $\frac{10}{15} = \frac{2}{3}$. এই যুক্তিটাই আমরা অংকের ফর্মুলা দিয়ে লিখব।

> (ii) Given that the student has failed in chemistry, the conditional probability that he has failed in mathematics is
>
> $$P(M|C) = \frac{P(M \cap C)}{P(C)} = \frac{0.1}{0.15} = \frac{2}{3}.$$

প্রথম অংশটাও একইভাবে হবে--

> (i) The required probability is
>
> $$P(M^c|C) = \frac{M^c \cap C}{P(C)} = \frac{0.05}{0.15} = \frac{1}{3}.$$

■

এই অংকগুলো খুব কঠিন কিছু নয়। কিন্তু এইরকম সহজ অংক দিয়েই এমন কাণ্ড করা যায় যে মাথায় গোল বেঁধে যাবে। এরকম একটা বিখ্যাত উদাহরণের কথা এখানে না বললেই নয়।

Example 4: আমাদের টিভিতে যেমন "কৌন বনেগা কড়োরপতি?"-জাতীয় রিয়েলিটি শো হয়, বিলেতে সেইরকম একটা জনপ্রিয় অনুষ্ঠানে একটা মজাদার জিনিস করা হত এইভাবে-- প্রথমে একজন দর্শককে মঞ্চে এনে তাঁকে তিনটে বন্ধ দরজা দেখানো হত। তার কোনো একটার পিছনে একটা দামী গাড়ি লুকোনো আছে, যেটা এই অনুষ্ঠানের পুরস্কার। এইটার লোভেই সেই দর্শক মঞ্চে উঠেছেন। অনুষ্ঠানপরিচালক এবার সেই দর্শককে বলবেন কোন দরজার পিছনে গাড়িটা আছে সেটা আন্দাজ করতে। আন্দাজ ঠিক হলে গাড়িটা তিনি পেয়ে যাবেন। তিনটে দরজাই একইরকম, তাদের যে কোনোটার পিছনেই গাড়িটা থাকতে পারে, সুতরাং আন্দাজটা নিতান্তই অন্ধের মত করা ছাড়া পথ নেই, এবং আন্দাজটা মিলে যাওয়ার probability হল $\frac{1}{3}$. খালি এইটুকু হলে এটা একটা মামূলী লটারী ছাড়া আর কিছুই হত না। কিন্তু ব্যাপারটার এখানেই শেষ নয়, আসল প্যাঁচটা আসছে এইবার। দর্শকের আন্দাজটা শোনার পর অনুষ্ঠানপরিচালক বাকি দুটো দরজার মধ্যে একটা খুলে দেবেন, যার পিছনে গাড়িটা নেই। কোন দরজাটা খুলবেন সেটা তিনি ঠিক করবেন এইভাবে--যদি দর্শকের আন্দাজটা ভুল হয়ে থাকে, তবে বাকি দরজা দুটোর মধ্যে ঠিক একটাই গাড়ি-বিহীন দরজা থাকবে। পরিচালক সেটাই খুলবেন। আর যদি দর্শক ঠিক দরজাটাই নির্বাচন করে থাকেন, তবে অবশিষ্ট দুটো দরজার মধ্যে যেকোনো একটা randomly খুলে দেখিয়ে দেওয়া হবে। মানে অনুষ্ঠানপরিচালক কোনোভাবেই গাড়িওয়ালা দরজাটা খুলে দেখাবেন না, সেটা বলাই বাহুল্য, এবং সে কথাটা দর্শকের অজানা নয়। এবার অনুষ্ঠানপরিচালক দর্শককে আরেকটা সুযোগ দেবেন--আপনি কি আপনার মত বদলাতে চান? রহস্যটা সেখানেই। মত কি বদলানো উচিত, নাকি নয়? ■

ভাবতে থাকো। আমরা উত্তরটা একটু পরে বলব। তার আগে একটা অংক করে নিই।

Example 5: If $P(ABC) = 0$, then show that $P(X|C) = P(A|C) + P(B|C)$, where $X = A + B$. (Notations as usual).[5] **(2011.1a)**

SOLUTION: এখানে ABC মানে $A \cap B \cap C$ আর $A + B$ মানে $A \cup B$. এখানে বাঁদিকে আছে $P(X|C)$, অর্থাৎ $\frac{P(X \cap C)}{P(C)}$. এখানে অবশ্যই ধরে নিতে হবে $P(C) > 0$ (যেটা বলে দেওয়া উচিত ছিল, নইলে conditional probability-গুলো অস্তিত্বই থাকে না)। প্রথমে $P(X \cap C)$ বার করা যাক--

We know

$$P(\underbrace{(A \cup B)}_{X} \cap C)$$

$$= P\big((A \cap C) \cup (B \cap C)\big)$$

$$= P(A \cap C) + P(B \cap C) - P\big((A \cap C) \cap (B \cap C)\big) \quad [\text{by incl. excl. principle}]$$

যেকোনো দুটো event যদি M, N হয়, তবে inclusion-exclusion principle বলে যে,
$P(M \cup N) = P(M) + P(N) - P(M \cap N)$.
সেটাই লাগালাম এখানে।

$$= P(A \cap C) + P(B \cap C) - P\big((A \cap B \cap C)$$

$$= P(A \cap C) + P(B \cap C) \quad [\because P(A \cap B \cap C) = 0.]$$

So, assuming $P(C) > 0$,

$$\begin{aligned} P(X|C) &= P(A \cup B|C) \\ &= \frac{P((A \cup B) \cap C)}{P(C)} \\ &= \frac{P(A \cap C) + P(B \cap C)}{P(C)} \\ &= \frac{P(A \cap C)}{P(C)} + \frac{P(B \cap C)}{P(C)} \\ &= P(A|C) + P(B|C), \end{aligned}$$

as required.

■

এবার আবার সেই গাড়ি জেতার প্রশ্নে ফিরে আসি। এটা নিয়ে এক সময়ে প্রচুর বিতর্ক চলেছিল। একদল বলেছিল যে সিদ্ধান্ত বদলিয়ে কোনো বাড়তি লাভ হবে না। তাদের যুক্তি এইরকম--

> অনুষ্ঠানপরিচালক জানে কোন দরজার পিছনে কী আছে। সে কখনোই গাড়িওয়ালা দরজাটা দেখাবে না, সে তো জানা কথাই। সেই জানা কথাটাই আরেকবার করে জেনে কী এমন বাড়তি তথ্য পাওয়া গেল যে সিদ্ধান্ত বদলাতে হবে?

যুক্তিটার মধ্যে খুব ভুল কিছু তো চোখে পড়ছে না। কিন্তু তাও আরেকদল বলল--

> গোড়ায় যখন সিদ্ধান্তটা নিয়েছিলাম তখন তো আমরা কিছুই জানতাম না কোন দরজার পিছনে কী আছে। এখন তো অন্ততঃ একটা দরজা জানলাম যার পিছনে গাড়িটা নেই। সেটা তো একটা নতুন তথ্য। তাহলে সেই তথ্যের আলোকে পুরোনো সিদ্ধান্তটাকে যাচাই করে নেওয়া উচিত নয় কী?

আরও পড়বার আগে ভেবে নাও তোমার কাছে কোন যুক্তিটা বেশী গ্রহণযোগ্য মনে হচ্ছে।
এবার আমরা সরাসরি অংক কষে আক্রমণ করি। এখানে random experiment-টা এইরকম--তিনটে দরজা আছে। এদের মধ্যে একটাকে randomly নির্বাচন করে তার পিছনে একটা গাড়ি রাখা হয়েছে। এখানে আমরা দুটো event নিয়ে মাথা ঘামাচ্ছি--

- $A = \{$দর্শকের নির্বাচিত দরজার পিছনে গাড়ি আছে$\}$,
- $B = \{$ অনুষ্ঠানপরিচালকের দেখানো দরজার পিছনে গাড়ি নেই$\}$.

আমাদের প্রশ্ন হল--কোনটা বড় $P(A|B)$ নাকি $P(A^c|B)$? মানে রেখো যে, আমরা জানি $P(A) = \frac{1}{3}$ এবং $P(B) = 1$. সংজ্ঞা লাগাই--

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)}.$$

শেষের ধাপটা পেলাম কারণ এখানে $P(A \cap B) = P(A)$. কেন বুঝে নাও--

> আমরা জানি যে, অনুষ্ঠানপরিচালক মরে গেলেও কখনো গাড়িওয়ালা দরজাটা দেখাবে না। তাই $P(B) = 1$ এবং $P(B^c) = 0$. এবার আমরা জানি যে,
>
> $$A = (A \cap B) \cup (A \cap B^c),$$

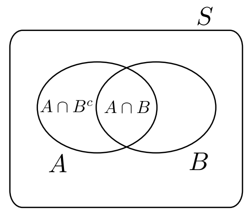


Fig 6

এবং ডানদিকের event দুটো disjoint. এখানে Fig 6 দেখে নিতে পারো। তাই

$$P(A) = P(A \cap B) + P(A \cap B^c).$$

এদিকে $P(A \cap B^c) \leq P(B^c) = 0$. সুতরাং $P(A \cap B^c) = 0$, কারণ probability তো আর negative হতে পারে না। তাই পাচ্ছি $P(A \cap B) = P(A)$.

যেহেতু $P(A) = \frac{1}{3}$ এবং $P(B) = 1$. সুতরাং $P(A|B) = \frac{1}{3}$. একইভাবে

$$P(A^c|B) = \frac{P(A^c \cap B)}{P(B)} = \frac{P(A^c)}{P(B)} = \frac{2/3}{1} = \frac{2}{3}.$$

সুতরাং আগের সিদ্ধান্ত বহাল রাখলে গাড়ি জেতার সম্ভাবনা $\frac{1}{3}$, আর সিদ্ধান্ত বদলালে probability একেবারে তার দ্বিগুণ, $\frac{2}{3}$. সুতরাং সিদ্ধান্ত বদলানোই বুদ্ধিমানের কাজ হবে।
আমরা জানি যে, $P(A) + P(A^c) = 1$ হয়। এটা probability-র axiom-গুলো থেকে প্রমাণ করা যায়। এই অংকে আমরা $P(A|B)$ আর $P(A^c|B)$ বার করলাম, ওদেরও যোগ করলে 1 হল। এমনটা কি যেকোনো A, B-এর জন্যই হবে, মানে যদি A, B যা খুশি দুটো event হয় (যেখানে $P(B) > 0$), তবে কি $P(A|B) + P(A^c|B) = 1$ হবেই? উত্তর হল--হ্যাঁ। শুধু তাই নয়, probability-র সব কয়টা axiom-ই conditional probability-র জন্যও খাটে। তারই প্রমাণ চেয়েছে নীচের অংকে।

Example 6: Define conditional probability, and show that such probabilities satisfies the axioms of probability. **(2006,2008)**

SOLUTION: প্রথম অংশ তো একটু আগেই করলাম।

Second part: Let B be some event with $P(B) > 0$.

Step 1: Shall show that for all events A we have $P(A|B) \geq 0$.

$$P(A|B) = \frac{P(A \cap B)}{B} \geq 0,$$

Since $P(A \cap B) \geq 0$ and $P(B) > 0$.

Step 2: Shall show that $P(S|B) = 1$, where S is the sample space.

$$P(S|B) = \frac{P(S \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Step 3: Shall show that for any countable collection of pairwise disjoint events $A_1, A_2, \ldots$ we have

$$P(\cup A_n|B) = \sum P(A_n|B).$$

This is because

$$\begin{aligned} P(\cup A_n|B) &= \frac{P((\cup A_n)\cap B)}{P(B)} \\ &= \frac{P(\cup(A_n\cap B))}{P(B)} \\ &= \frac{\sum P(A_n\cap B)}{P(B)} \quad [\because A_n\cap B\text{'s are disjoint, since } A_n\text{'s are.}] \\ &= \sum \frac{P(A_n\cap B)}{P(B)} \\ &= \sum P(A_n|B), \end{aligned}$$

as required.

■

Example 7: Prove that $P(B|A) \geq 1 - P(\bar{B})/P(A),\ P(A) > 0.$[5] **(2012, 2007, 2004)**

SOLUTION: দেওয়া আছে $P(A) > 0$. আর দেখাতে হবে $P(B|A) \geq 1 - P(\bar{B})/P(A)$.

$$\begin{aligned} P(B|A) &= \frac{P(A\cap B)}{P(A)} \\ &= \frac{P(A) - P(A\cap B^c)}{P(A)} \\ &\quad [\because A = (A\cap B)\cup(A\cap B^c) \text{ and } (A\cap B)\cap(A\cap B^c) = \phi, \\ &\quad \therefore P(A) = P(A\cap B) + P(A\cap B^c).] \\ &= 1 - \frac{P(A\cap B^c)}{P(A)} \\ &\geq 1 - \frac{P(B^c)}{P(A)}, \end{aligned}$$

as reqd, since $(A\cap B^c) \subseteq B^c$, and so $P(A\cap B^c) \leq P(B^c)$.

■

Example 8: A and B are two independent witnesses in a case. The probability that A will

speak the truth is x, and the probability that B will speak the truth is y. A and B agree in a certain statement. Show that the probability that this statement is true is

$$\frac{xy}{1-x-y+2xy}.$$

(2009.2d)

SOLUTION: এই অংকটা একেবারেই ঠিক করে লেখেনি। একটা গুরুত্বপূর্ণ কথা বাদ দিয়ে গেছে--

> We assume that the issue is of yes/no type, where both A and B know the truth. Thus if they both lie, then their statements agree.

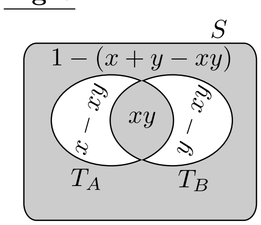
অর্থাৎ কিনা প্রশ্নটা হ্যাঁ-না টাইপের। তাই যদি আসল উত্তরটা হয় "হ্যাঁ" (যেটা দুই সাক্ষীই জানে) আর দুজনেই মিথ্যে বলে, তবে দুজনেই "না" বলবে। একইভাবে আসল উত্তরটা "না" হলে দুজনেই "হ্যাঁ" বলবে। যদি উত্তরটা হ্যাঁ-না টাইপের না হত, তবে দুজনে দুরকম মিথ্যে বলে ফেলতে পারত। এই প্রসঙ্গে একটা মজার গল্প মনে পড়ল। দুটো ছাত্র পরীক্ষার আগের দিন শহরে গিয়ে খুব হুল্লোড় করে রাতে আর কলেজে ফিরতে পারে নি। তাই পরীক্ষাও দিতে পারে নি। পর দিন কাঁচুমাচু মুখে প্রফেসরকে গিয়ে বলেছে--"স্যার, আসলে কী হয়েছিল জানেন তো, আমাদের পরিচিত এক দরিদ্র ভদ্রলোকের খুব শরীর খারাপ, তাঁকে শুশ্রূষা করার জন্য আমরা শহরে গিয়েছিলাম, ফেরার পথে গাড়ির একটা টায়ার গেল ফেটে, তাই সারা রাত রাস্তাতেই কাটাতে হয়েছিল। তাই যদি স্যার আমাদের পরীক্ষাটা আরেকদিন একটু নিয়ে নেন।" প্রফেসর ভালো মানুষের মত মুখ করে বললেন--তা বেশ তো, তোমরা কালকেই পরীক্ষা দাও না কেন! আমি নতুন করে প্রশ্ন বানিয়ে দেব।" পর দিন দুই শ্রীমান পরীক্ষা দিতে গিয়ে দেখে দুজনের আলাদা আলাদা ঘরে সিট দেওয়া হয়েছে। যথাসময়ে পরীক্ষা শুরু হল। প্রশ্নপত্র খুলে তাদের চক্ষুস্থির! খালি একটাই প্রশ্ন আছে পেপারে--চারটে টায়ারের মধ্যে ঠিক কোন টায়ারটা ফেটেছিল লেখো।

আবার আমাদের অংকে ফিরে আসি। প্রশ্নে কী চেয়েছে সেটা বুঝে নাও। লিখেছে "probability that this statement is true". অর্থাৎ A, B দুজনেই সত্য বলেছে। সবকিছু অংকের ভাষায় প্রকাশ করে নিই--

> Let T_A (resp, T_B) be the event that A (resp, B) speaks the truth. Then their statements agree if both of them speak the truth or if both of them lie. This event is
>
> $$M = (T_A \cap T_B) \cup (T_A^c \cap T_B^c).$$

এইটা আমাদেরকে দেওয়া আছে। Fig 7-এ এটাকে দেখিয়েছি ছাইছাই রঙে। ছবিতে x, y ইত্যাদিও probability-গুলোও দেখিয়েছি। সেগুলোর কথায় একটু পরে আসছি।

Fig 7



The event that both of them speak the truth is

$$N = T_A \cap T_B.$$

এইটার conditional probability আমাদের বার করতে হবে। Fig 7-এ এটা হল মাঝের খাড়া চোখের মত অংশটা।

Now

$$\begin{aligned} P(N|M) &= \frac{P(N \cap M)}{P(M)} \\ &= \frac{P(N)}{P(M)} \quad [\because N \subseteq M] \end{aligned}$$

এবার আমরা $P(N)$ আর $P(M)$ বার করব।

Now since T_A, T_B are independent, $P(N) = P(T_A)P(T_B) = xy$. Also,

$$\begin{aligned} P(M) &= P(T_A \cap T_B) + P(T_A^c \cap T_B^c) \quad [\because (T_A \cap T_B) \text{ and } (T_A^c \cap T_B^c) \text{ are disjoint }] \\ &= P(T_A)P(T_B) + P(T_A^c)P(T_B^c) \end{aligned}$$

$[\because T_A, T_B$ are independent, so are T_A^c, T_B^c.]
দুটো event যদি independent হয়, তবে তাদের
complement-রাও যে independent হয়,
সেটা আমরা প্রথম অধ্যায়েই প্রমাণ করেছিলাম।

$$= xy + (1-x)(1-y) = 1 - x - y + 2xy.$$

এবার খালি conditional probability-টা বার করার অপেক্ষা--

$\therefore P(N|M) = \frac{P(N)}{P(M)} = \frac{xy}{1-x-y+2xy}$, as required.

■

10.1 Independence-এর সঙ্গে সম্পর্ক

আমরা প্রথম অধ্যায়েই independence-এর কথা শিখেছি। সহজ ভাষায় মানেটা ছিল এই--যদি A আর B দুটো event হয়, তবে "ওরা independent" মানে ওদের একজনকে জানলে অন্যজনের সম্পর্কে কোনো তথ্য পাওয়া যায় না। যেমন ধরো একটা কয়েন আছে যার $P(\text{head}) = \frac{1}{2}$. পরপর দুবার একটা কয়েন টস করা হলে প্রথম টসের ফলাফল থেকে দ্বিতীয় টসের ফলাফল কোনোভাবেই আন্দাজ করা সম্ভব নয়। অংকের ভাষায় সংজ্ঞাটা ছিল $P(A \cap B) = P(A)P(B)$.
আমরা অনেক সময়ে independence-এর সংজ্ঞাটাকে চট্ করে ভেবে ফেলি এইভাবে--A আর B-এর মধ্যে কোনো সম্পর্ক নেই। সেটা যে সঠিক নয়, সেটা conditional probability দিয়ে চিন্তা করলে বোঝা যাবে। একটা উদাহরণ নিই।

Example 9: এই অধ্যায়ের গোড়াতেই আমরা একটা ক্যাসিনোর উদাহরণ বলেছিলাম, যেখানে একটা বাক্সের মধ্যে একটা ঘড়ির মত লেখা সংখ্যা ছিল। এটা সেই উদাহরণেরই রকমফের। বাক্সের মধ্যে যে সংখ্যাটা ছিল, সেটা $0, 1, 2, ..., 9$ যা খুশি হতে পারত, সবারই সমান probability. তোমাকে বার করতে হচ্ছিল $P(A)$, যেখানে A মানে "সংখ্যাটা জোড়"। বোঝাই যাচ্ছে উত্তর হল $P(A) = \frac{1}{2}$. এই পর্যন্ত আগের মতই। খালি ওখানে দরজায় একটা ফুটো ছিল, যা দিয়ে তুমি দেখে

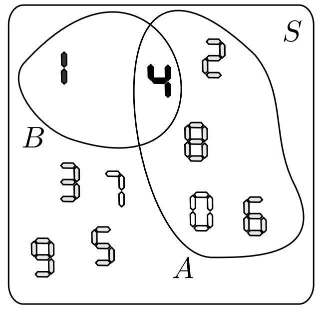


Fig 8

ফেলেছিলে যে, মাঝের আলোটা জ্বলছে না। এবারও ফুটো আছে, তবে সেটা একটু ওপরে। তা দিয়ে খালি দেখা যাচ্ছে যে, একেবারে ওপরের আলোটা নিভে আছে। এবার A-র conditional probability-টা কী হবে?

SOLUTION: চট্ করে ভেবে নাও, কোন কোন সংখ্যার জন্য ওপরের আলোটা নিভে থাকে (Fig 8)। উত্তর হল--খালি 1 আর 4-এর জন্য। এদের মধ্যে ঠিক একজনই জোড়। অতএব $P(A|B) = \frac{1}{2}$. কিন্তু $P(A)$ তো সেই $\frac{1}{2}$-ই ছিল, তাহলে B ঘটেছে জেনে লাভ কী হল? এক্ষেত্রে লাভ কিছুই হল না। তাই আমরা বলব যে, A আর B এখানে independent. এখানে

$$P(A|B) = P(A),$$

অর্থাৎ

$$\frac{P(A \cap B)}{P(B)} = P(A),$$

মানে

$$P(A \cap B) = P(A)P(B).$$

কিন্তু এখানে A আর B মধ্যে "কোনো সম্পর্ক নেই" এমনটা কিন্তু মোটেই মনে হচ্ছে না। B হয়েছে জানার ফলে আমাদের sample space-টা $\{0, 1, 2, ..., 9\}$ থেকে ধাঁ করে কমে একেবারে $\{1, 4\}$ হয়ে গেল, সুতরাং B-এর মধ্যে তথ্য যে আছে সেটা অনস্বীকার্য, কিন্তু তাতে $P(A)$ থেকে $P(A|B)$ বদলাচ্ছে না। এটাই হল independence-এর মর্ম। ■

এই আলোচনাটা বুঝে থাকলে নীচের ছোট্টো অংকটা খুবই সহজ।

Exercise 1: যদি $P(A|B) = P(A)$ হয়, তবে দেখাও যে $P(B|A) = P(B)$ হবে। এখানে $P(A), P(B) > 0$. অর্থাৎ, A জানলে যদি B-এর probability না বদলায়, তবে B জানলে A-র probability-ও বদলাবে না। ■

DAY 11
Bayes' theorem

আমরা ছোটোবেলায় অংক শেখার সময়ে সাধারণতঃ প্রথমে arithmetic (পাটীগণিত) এবং পরে algebra (বীজগণিত) শিখি। একটা arithmetic-এর অংক ধরো এইরকম--

> 7-এর সঙ্গে 3 যোগ করলে কত হয়?

উত্তর অবশ্যই 10. এটা নীচু ক্লাসেই শেখা হয়। একটু উঁচু ক্লাসে উঠে এর উল্টো অংকটা শেখানো হয়--

> একটা সংখ্যা আছে যার সঙ্গে 3 যোগ করলে 10 হয়। বলো তো সংখ্যাটা কী?

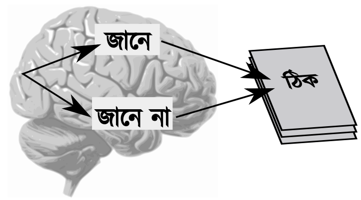


Fig 9

এর সমাধান করা হয় algebra দিয়ে-- প্রথমে অজানা সংখ্যাটাকে x ধরে নিয়ে একটা equation পাওয়া যায় $x+3=10$, সেটাকে সমাধান করার জন্য algebra-র একটা "ম্যাজিক কায়দা" আছে (যেটা ছোটোবেলায় অনেকেই না বুঝে মুখস্থ করে) x-টাকে বাঁদিকে রেখে, বাকী সব ডানদিকে নিয়ে যেতে হয়, এবং বাঁদিকে যেটা ছিল $+3$, সেটাই ডানদিকে যাবার পথে কী করে যেন -3 হয়ে যায়, মানে $x = 10-3 = 7$. এই রকম উল্টো দিকে যাওয়াটা অংকের দুনিয়ায় সর্বত্রই দেখা যায়। এবং প্রায়শঃই দেখা যায় যে, যে জিনিসটা সোজা দিকে নিতান্তই সহজ, সেটাই উল্টো দিক দিয়ে দেখলে রীতিমত কঠিন হয়ে পড়ে। যেমন, কোনো সংখ্যাকে নিজের সঙ্গে গুণ করে square বার করা নিতান্তই সহজ। কিন্তু তা বলে একটা square number দিলে তার square root বার করা তত সহজ নয়। এইবার এরকম একটা ব্যাপার দেখব probability নিয়ে। একটা উদাহরণ নিয়ে শুরু করলে সুবিধা হবে।

Example 10: আজকাল MCQ-এর খুব চল হয়েছে। মাষ্টারমশাইরা এতে খুশি, কারণ খাতা দেখার ঝকমারি কম। আবার

ছাত্রছাত্রীরাও খুশি, কারণ আন্দাজে মেরে দেবার সুযোগটা বেশী। মনে করো একজন মাষ্টারমশাই একটা MCQ দেখছেন। অনেক ছাত্র পরীক্ষা দিয়েছে। তাদের কেউ উত্তরটা বার করতে জানে, কেউ বা জানে না। কে জানে আর কে জানে না সেটা আগে থেকে মাষ্টারমশাইয়ের তো আর জানা নেই। তাই তাঁর কাছে ব্যাপারটা ঠিক কয়েন টসেরই মত--"head" মানে 'জানে" আর "tail" মানে "জানে না"। মনে হতে পারে মাষ্টারমশাই ঠিক উত্তর দেখলেই নম্বর দিয়ে দেবেন, আর ভুল উত্তর দেখলে নম্বর দেবেন না। কিন্তু এখানে একটা সমস্যা আছে। যদি তিনি দ্যাখেন যে, একজন ঠিক উত্তর দিয়েছে, তার মানেই কিন্তু এই নয় যে, ছাত্রটা সেটা সত্যিই জেনেবুঝে দিয়েছে। হতে পারে স্রেফ আন্দাজে লাগিয়ে দিয়েছে। সুতরাং এখানেই সেই উল্টো পথে যাওয়ার প্রয়োজন। ছাত্রটা "সত্যিই কতটা জানে" তার সঙ্গে "আন্দাজ" মিলে যেটা তৈরী হয়েছে সেটাই হল "লিখিত উত্তর"। মাষ্টারমশাইয়ের কাজ হল ওই "আন্দাজ" অংশটুকু ছেঁকে ফেলে "সত্যিই কতটা জানে" সেটা বার করে আনা। ব্যাপারটা এইভাবে লেখা যায়--

"সত্যিই কতটা জানে" $+$ "আন্দাজ" $=$ "লিখিত উত্তর"।

তা থেকে মাষ্টারমশাইয়ের কাজ হল উল্টে এইটা বার করা--

"সত্যিই কতটা জানে" $=$ "লিখিত উত্তর" $-$ "আন্দাজ"

বুঝতেই পারছ যে, এখানে '$+$' বা '$-$' চিহ্নগুলো নেহাতই মজা করে বসানো। এখানে আমরা মোটেই সংখ্যার যোগ বিয়োগ করছি না। এখানে তাই আলাদা হাতিয়ার দরকার। এবং সেই হাতিয়ারই হল Bayes' theorem, যেটা আমরা এবার শিখব। ■

Bayes' theorem শেখার আগে উচ্চারণটা শিখে রাখি। Bayes-এর উচ্চারণ হল "বেইজ", খালি 'জ'-টাকে Z-এর মত করে বলতে হবে। কাউকে কাউকে পায়েসের মত করে "বায়েস" বলতে শুনেছি। সে উচ্চারণটা মোটেই ঠিক নয়।
Bayes' theorem বোঝার জন্য আবার আমরা MCQ-এর সমস্যায় ফিরে যাই। ব্যাপারটাকে Fig 9-এর মত ছবি দিয়ে ভাবা যায়--সত্যিই জানে নাকি জানে না, সেটা ছাত্রের মগজের ভিতরের ব্যাপার। সেটা মাষ্টারমশায় দেখতে পাচ্ছেন না। তিনি যেটা দেখতে পাচ্ছেন সেটা হল লিখিত উত্তরটা। এখানে "ঠিক" উত্তরে দুই পথে পৌঁছনো যাচ্ছে--

- জানে, তাই ঠিক উত্তর

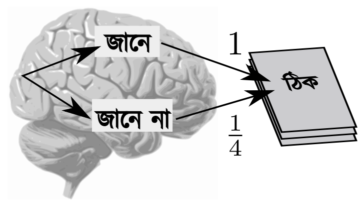


Fig 10

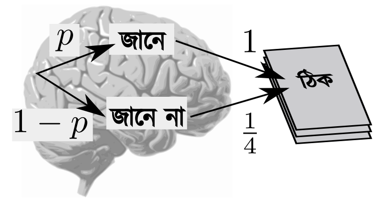


Fig 11

- জানে না, কিন্তু আন্দাজে ঠিক উত্তর

মাষ্টারমশায়ের আগ্রহ এদের মধ্যে প্রথম পথটা নিয়ে। এর জন্য আমরা একটা কৌশল করব। মনোযোগ দিয়ে পড়। যদি কেউ উত্তরটা আন্দাজে দিয়ে থাকে তবে ঠিক হবার probability কত? ধরো MCQ-তে চারটে option ছিল, যাদের মধ্যে ঠিক একটাই ঠিক। যদি ছাত্রটা একেবারে আন্দাজে ঢিল ছুঁড়ে থাকে, তাহলে ঠিক হবার probability হবে $\frac{1}{4}$. এ থেকে কি আমরা বলতে পারি $P(\text{ঠিক উত্তর}) = \frac{1}{4}$? না, কারণ এটা খালি আন্দাজের কেসটা ধরে হিসেব করা। তাই আমরা লিখব

$$P(\text{ঠিক} \mid \text{জানে না}) = \frac{1}{4}.$$

ছবিতে "জানে না" থেকে "ঠিক" উত্তর পর্যন্ত যে তীরচিহ্নটা আছে, তার পাশে লিখব $\frac{1}{4}$. এবার কোনো অংক না কষে বলো তো-- "জানে" থেকে "ঠিক" উত্তরে পৌঁছোনোর probability কত? অবশ্যই 1, কারণ এক্ষেত্রে আন্দাজের কোনো প্রশ্নই নেই। সুতরাং পেলাম Fig 10। এবার প্রথম দুটো তীর চিহ্নের দিকে তাকাও। ওদের জন্য এবার probability চাই। এটা নির্ভর করছে প্রশ্নটা কত কঠিন তার উপর, যদি খুব কঠিন হয়, তবে খুব কম ছাত্রই উত্তরটা জানবে, ধরো মোটে 10% জানে। সেক্ষেত্রে "জানে" পর্যন্ত যে তীর চিহ্নটা, তার পাশে 0.1 বসবে, আর "জানে না"-র জন্য 0.9. আবার যদি খুব সহজ প্রশ্ন হয়, তবে হয়তো 99% ছাত্রই উত্তরটা জানবে, সেক্ষেত্রে "জানে" আর "জানে না"-র জন্য যথাক্রমে 0.99 আর 0.01 বসবে। আমরা কাজটা in general করব, p আর $1-p$ ধরে নিয়ে। তাহলে পেলাম Fig 11 এখানে উপরের পথ ধরে "ঠিক" উত্তরে পৌঁছবার probability হবে $p \times 1$, মানে যে যে তীর ধরে যাচ্ছি তাদের probability-গুলোর গুণফল। একইভাবে নীচের পথটার probability হবে $(1-p) \times \frac{1}{4}$. সুতরাং "ঠিক" উত্তরের probability পাওয়া যাবে এ দুই পথের probability যোগ করলে--

$$P(\text{ঠিক}) = p \times 1 + (1-p) \times \tfrac{1}{4} = \tfrac{1}{4} + \tfrac{3}{4}p.$$

লক্ষ কর যে, এর মধ্যে উপরের পথটার অবদান হল $p \times 1$. সুতরাং একজন ছাত্র ঠিক উত্তর দিয়ে থাকলে, তার সত্যি সত্যি জেনেবুঝে উত্তর দেওয়ার conditional probability হল

$$P(\text{জানে} \mid \text{ঠিক}) = \frac{p}{\frac{1}{4} + \frac{3}{4}p}.$$

এইটা কি একটু গুলিয়ে গেল? তাহলে হয়তো অংকের ভাষায় লিখলে সুবিধা হবে। এর জন্য খালি এই কয়টা জিনিস মনে রাখতে হবে--

1. যদি A, B দুটো event হয় যাতে $P(A), P(B) > 0$ তবে $P(A \cap B)$-কে দুভাবে লেখা যায়--

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B).$$

2. যদি A, B যেকোনো দুটো event হয়, তবে B-কে দুটো disjoint subset-এ ভেঙে লেখা যায়--

$$B = (A \cap B) \cup (A^c \cap B).$$

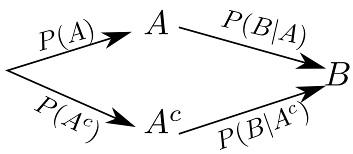


Fig 12

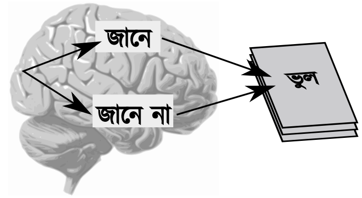


Fig 13

এবার আমাদের উদাহরণে আমরা $A =$"জানে" এবং $B =$"ঠিক উত্তর" নেব। সুতরাং $A^c =$"জানে না" হবে, বুঝতেই পারছ। এবার ধাপে ধাপে এগোই--

1. ধরো একজন ঠিক উত্তর দিয়েছে, মানে B ঘটেছে। সেটা দুভাবে হতে পারে, জেনেবুঝে বা না জেনে--
$$B = (A \cap B) \cup (A^c \cap B).$$

2. সুতরাং probability নিলে হবে
$$P(B) = P(A \cap B) + P(A^c \cap B).$$
এখানে ডানদিকে দুটো probability যোগ হল, কারণ event দুটো disjoint ছিল, ("জানে' এবং "জানে না" দুটোই তো আর একই সঙ্গে হতে পারে না!))।

3. আমরা জানি $P(A \cap B) = P(A)P(B|A)$ আর $P(A^c \cap B) = P(A^c)P(B|A^c)$.

4. সুতরাং পেলাম
$$P(B) = P(A)P(B|A) + P(A^c)P(B|A^c).$$

5. তোমায় বার করতে হবে $P(A|B)$. সেটা হল $P(A \cap B)/P(B)$. অতএব
$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}.$$

এইবার এই পুরো জিনিসটা তীর চিহ্ন দিয়ে যা পেয়েছিলাম তার সঙ্গে মিলিয়ে নাও। সুবিধার জন্য Fig 12 ব্যবহার করতে পারো।
এই পুরো জিনিসটা কিন্তু একবারে মাথায় না ঢোকাই স্বাভাবিক। কয়েকবার করে পড়ে নাও। তারপর দ্যাখো তো নীচের অংকটা করতে পারছ কিনা।

Exercise 2: আবার সেই MCQ নিয়েই কাজ করছি। তবে এবার যে ছাত্রের খাতা দেখা হচ্ছে সে অংকটা ভুল করেছে (Fig 13)। এই তথ্যটা শোনার পর এখন বলতে হবে তার উত্তরটা জানার conditional probability কত, মানে P(জানে | ভুল উত্তর) বার করতে চাই। আশা করি বুঝতেই পারছ যে এটা হওয়া উচিত 0, কারণ জানলে আর ভুল উত্তর দিল কেন? (এখানে আমরা "আহা, বোধহয় নার্ভাস হয়ে গেছিল", কিংবা "ইশ্‌শ্‌, ভুল জায়গায় টিক দিয়ে ফেলেছে" ইত্যাদি সম্ভাবনা ধরছি না)। আমরা যে দুটো কায়দা শিখলাম (এক, তীর চিহ্ন দিয়ে, আর দুই, অংক কষে) দুভাবেই করে দ্যাখো তো উত্তরটা 0-ই পাও কিনা। ■

তোমরা জানো যে অনেক MCQ-তেই negative marking থাকে। তার পিছনে যুক্তি এরকমই। নম্বরটা কেবল "ঠিক" আর "ভুল"-এর ভিত্তিতে না দিয়ে P(জানে | ঠিক) আর P(জানে | ভুল)-এর ভিত্তিতে দেওয়া উচিত। অর্থাৎ--

ঠিক উত্তর হলেই 1 এবং ভুল উত্তর হলেই 0

এভাবে না দিয়ে



Fig 14

ঠিক উত্তর হলেই $\frac{4p}{1+3p}$ এবং ভুল উত্তর হলে 0

এইভাবে দেওয়াটা বেশী ভালো। লক্ষ কর এখানে ঠিক উত্তরে নম্বরটা একটু কমিয়ে দেওয়া হয়েছে, আন্দাজ করার ব্যাপারটা কাটানোর জন্য। কতটা কমানো হয়েছে সেটা p-এর উপর নির্ভর করছে, মানে প্রশ্নটা কতটা সহজ ছিল তার উপরে। যদি p বড় হয় (মানে সহজ প্রশ্ন), তবে এই নম্বরটা বাড়বে, যদি ছোটো হয় (মানে কঠিন প্রশ্ন) তবে নম্বরটাও কমবে। যদি একটা কঠিন প্রশ্নে ঠিক উত্তর দাও, তবে আন্দাজের সম্ভাবনাটা বেশী, তাই নম্বরটা কম পাচ্ছ। আর যদি সহজ প্রশ্নে ঠিক উত্তর দাও, সেক্ষেত্রে আন্দাজের সম্ভাবনাটা কম, তাই নম্বরটা বেশী পাচ্ছ। অংক কষে যতই প্রমাণ করা যাক, দুঃখের ব্যাপার হল অধিকাংশ ছাত্রই এটা ঠিক মেনে নিতে পারে না। ভুল করলে 0 পাচ্ছি সেটা বুঝলাম, কিন্তু ঠিক করলেও কেন নম্বর কাটা যাবে? তাও আবার বেশী কঠিন প্রশ্ন ঠিক করলে বেশী নম্বর কাটা যাচ্ছে!

--"কে কোথায় আন্দাজে মেরে দেবার চেষ্টা করেছে বলে আমার নম্বর কেটে দেবে? মামাবাড়ি পেয়েছে?"

সেই ক্রুদ্ধ জনতার কথা ভেবেই একই জিনিসকে একটু মনমত করে পরিবেশন করা হয়--ঠিক উত্তরে পুরো নম্বর, আর ভুল উত্তরে negative marking. প্রশ্ন যত সহজ হয় negative marking-এর পরিমাণও তত বাড়ে। ব্যাপারটা এইরকম--

- হুম্‌ম্‌, ভুল করেছে বটে, কিন্তু অংকটা বেশ কঠিনই ছিল, বেচারীকে খুব বেশী দোষ দেওয়া যায় না (কম negative marking)।
- সে কি, এই সহজ প্রশ্নটাও ভুল করেছে! ধুস্‌ এই ছাত্রটাকে দিয়ে কিস্যু হবে না (বেশী negative marking)।

অবশ্য সাধারণতঃ যাঁরা MCQ-এর প্রশ্ন বানান তাঁরা সবগুলো প্রশ্নকেই একইরকম কঠিন রাখেন। তাই সব প্রশ্নেই একই রকম negative marking ব্যবহার করা যায়। শুনেছি GRE ইত্যাদি বিদেশী কিছু পরীক্ষায় প্রতিটা প্রশ্ন কতটা কঠিন সেটা আলাদা আলাদা করে হিসেবে নেওয়া হয়।
এই যে MCQ নিয়ে এত আলোচনা করলাম, এর মূল নির্যাসটুকু কিন্তু অন্য বহুক্ষেত্রেও লাগে। বিজ্ঞানের যেকোনো শাখাতেই বিভিন্ন জিনিসের পরিমাপ করতে হয়, এই সব জিনিসের অনেকগুলোই আসলে random (অর্থাৎ তাদের পিছনের সম্পূর্ণ কার্যকারণ বৈজ্ঞানিকের কাছে অজানা)।

Example 11: যেমন ধরো, একটা বনে কতগুলো গাছ আছে বার করতে হবে। বিভিন্ন বনে বিভিন্ন সংখ্যক গাছ থাকতে

পারে, সেটা আগে থেকে নিশ্চিত হয়ে জানা সম্ভব নয়। সুতরাং সেখানে একপ্রস্থ randomness রয়েছে। এবার যখন তুমি কোনো একটা বনের গাছের সংখ্যা গুণবে, তখন তো আর ধরে ধরে প্রত্যেকটা গাছকে গোণা সম্ভব নয়, তাই আমরা বনের ম্যাপের উপরে গ্রাফকাগজের মত একটা ছক এঁকে নিই। Fig 14 দ্যাখো[2]। প্রত্যেকটা চৌকোর area-কে ধরো 1 বললাম। এবার কোনো একটা random চৌকো নিয়ে সেইখানকার গাছগুলোকে ধরে ধরে গোণা হয়। এবার বনের মোট area দিয়ে গুণ করলে বনের মোট গাছের সংখ্যার একটা আন্দাজ পাওয়া যাবে। যেমন Fig 14-এর একটা random চৌকো মোটা করে দেখিয়েছি। এর মধ্যে রয়েছে 3-খানা গাছ। মোট চৌকোর সংখ্যা হল 53. সুতরাং গুণফল হচ্ছে $3 \times 53 = 159$. এই উত্তরটা

[2]এখানে গাছগুলোকে ছোটো ছোটো বিন্দু দিয়ে দেখিয়েছি। সত্যিকারের ম্যাপে অবশ্যই প্রত্যেকটা গাছ দেখানো থাকে না। সেরকম থাকলে তো ম্যাপ দেখেই গুণে ফেলা যেত।

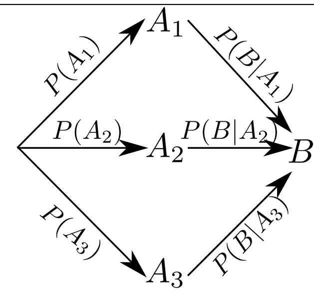


Fig 15

হল মোট সংখ্যার একটা approximation, কারণ আমরা কোন চৌকোটায় ধরে ধরে গুণেছিলাম, সেটা ছিল random. এইখানে আরেকপ্রস্থ randomness ঢুকল। ■

এইভাবে অনেক জিনিসকেই মাপতে গেলেই দুই প্রস্থ randomness ঢুকে পড়ে, প্রথম প্রস্থটা প্রকৃতির নিজস্ব randomness, আর দ্বিতীয়টা হল আমাদের পরিমাপপদ্ধতির অসম্পূর্ণতায়।

Example 12: MCQ-এর উদাহরণে যেমন ছাত্রদের "জানা বা না-জানা"-টা ছিল প্রাকৃতিক randomness. পরিমাপপদ্ধতিটা ছিল পরীক্ষা নেওয়া, যেখানে আন্দাজে মেরে দেওয়ার সুযোগ রয়েছে। সেটা হল দ্বিতীয় দফার randomness. এরকম আরও উদাহরণ দেওয়া যায়--টিউমার দেখে বলা ক্যান্সার হয়েছে কিনা। সত্যিই ক্যান্সার হয়েছে কিনা তার একটা নিজস্ব probability আছে, সেটা আসছে প্রাকৃতিক randomness থেকে। এবার ডাক্তারবাবু যে সব পরীক্ষা-নিরীক্ষা করে সিদ্ধান্ত করছেন সেগুলো একেবারে নিখুঁত নয়, ভুলভ্রান্তি হতেই পারে। সেখানেই দ্বিতীয় প্রস্থ randomness. ■

এরকম ক্ষেত্রে প্রাকৃতিক randomness-টা নিয়েই বৈজ্ঞানিকের আগ্রহ থাকে, পরিমাপের randomness-টা উট্‌কো ঝামেলামাত্র। তাই সেই randomness-টা "কাটিয়ে" খালি প্রাকৃতিক randomness-টাকে বার করে আনতে পারলে ভালো হয়। সেই কাজেই Bayes' theorem-এর বহুল ব্যবহার। বস্তুতঃ এটা এতই জনপ্রিয় প্রয়োগ যে, এর ভিত্তিতে statistics-এর একটা আস্ত শাখাই গড়ে উঠেছে, যার নাম Bayesian statistics.
যাই হোক, আমরা এই বইতে অতসবের মধ্যে যাব না। আমরা খালি এই যুক্তিটাকেই গুছিয়ে একটা theorem-এর আকার দেব, যার নাম Bayes' theorem. এই theorem-টাকে নানাভাবে লেখা যায়। সবগুলোরই মূল ব্যাপার একই--

> দুই প্রস্থ randomness থাকবে। এই দুই প্রস্থ randomness-এর পর একেবারে শেষে কী ঘটেছে সেটা জানা আছে। তার ভিত্তিতে তোমাকে বার করতে হবে প্রথম প্রস্থ randomness-এর পরে কী ঘটেছিল।

আমরা যে সংস্করণটা এই বইতে লিখে প্রমাণ করব তার জন্য Fig 15 দ্যাখো। এখানে প্রথম প্রস্থ randomness-এর পরে n রকমের ঘটনা হতে পারে, যাদেরকে $A_1, ..., A_n$ দিয়ে দেখিয়েছি। (ছবিতে $n = 3$ নিয়েছি)। MCQ-এর উদাহরণে $n = 2$ ছিল, A_1 মানে "জানে", A_2 মানে "জানে না"। এই A_i-রা ছাড়া আর কিছু হতে পারে না (অর্থাৎ A_i-রা mutually exhaustive) এবং দুটো A_i কখনও একই সঙ্গে ঘটতে পারে না (অর্থাৎ A_i-রা pairwise exclusive). দুই প্রস্থ randomness-এর পর যেটা ঘটেছে বলে জানা আছে, সেটা হল B. আমাদের MCQ-এর উদাহরণে B ছিল "ঠিক" উত্তর। এইবার ঠিক আগের মতই এগোতে হবে। B-তে পৌঁছনোর n-টা পথ আছে। তাদের মধ্যে কোন পথটা নেওয়া হয়েছে সেইটা নিয়েই মাথা ঘামাব। যেমন ধরো, Fig 15-এ সবচেয়ে উপরের পথের probability হল

$$P(A_1) \times P(B|A_1).$$

এইটা পাওয়া গেল সেই পথ বরাবর তীর দুটোর probability গুণ করে। মোট তিন পথে B-তে পৌঁছনো যাচ্ছে। এদের probability-গুলো যোগ করে পাব--

$$P(B) = P(A_1) \times P(B|A_1) + P(A_2) \times P(B|A_2) + P(A_3) \times P(B|A_3).$$

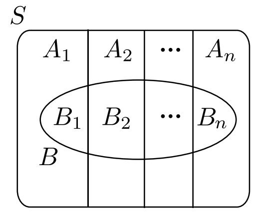


Fig 16

সুতরাং $P(A_1|B)$ হবে

$$P(A_1|B) = \frac{P(A_1) \times P(B|A_1)}{P(A_1) \times P(B|A_1) + P(A_2) \times P(B|A_2) + P(A_3) \times P(B|A_3)}.$$

এই ব্যাপারটা বুঝে নিয়ে এবার নীচের অংকের সমাধানটা পড়।

Example 13: State and prove Bayes' theorem.[5] **(2013,2011,2009,2005)**

SOLUTION:

Bayes' theorem

Let $A_1, ..., A_n$ be mutually exhaustive and pairwise disjoint events, such that $\forall i \quad P(A_i) > 0$. Let B be any event with $P(B) > 0$. Then

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}.$$

Proof: Let S be the sample space. Then

$$S = A_1 \cup \cdots \cup A_n.$$

Fig 16-এর দিকে চোখ রাখো।

So

$$B = B \cap S = B \cap (A_1 \cup \cdots \cup A_n) = \underbrace{(B \cap A_1)}_{B_1} \cup \cdots \cup \underbrace{(B \cap A_n)}_{B_n}.$$

Since $\forall j \quad B_j \subseteq A_j$, and A_j's are pairwise disjoint, so B_j's are also pairwise disjoint.

Hence

$$P(B) = \sum_{j=1}^{n} P(B_j).$$

Now $P(B_j) = P(B \cap A_j) = P(A_j)P(B|A_j)$, by definition of conditional prob-

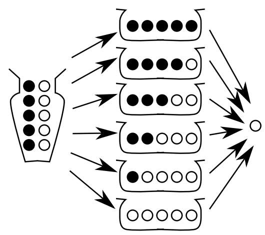


Fig 17

> ability.
> $\therefore P(B) = \sum_j P(B_j) = \sum_j P(A_j)P(B|A_j)$.
> So
>
> $$P(A_i|B) = \frac{P(B \cap A_i)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)},$$
>
> as required.

■

এবার একটা প্রয়োগ দেখা যাক।

Example 14: From an urn containing 5 white and 5 black balls, 5 balls are transferred at random into an empty second urn from which one ball is drawn, and it is found to be white. What is the probability that all balls transferred from the first urn are white?[5] **(2011.1b)**

SOLUTION: এখানে urn (আর্ন) মানে হল একধরণের পাত্র। এরকম দুটো পাত্র নিয়ে আমরা শুরু করছি, প্রথমটায় আছে 5-টা সাদা আর 5-টা কালো বল। দ্বিতীয় পাত্রটা আপাততঃ ফাঁকা। প্রথম পাত্র থেকে 5-টা বল তোলা হল random-ভাবে, এবং দ্বিতীয় পাত্রে রাখা হল। এই হল প্রথম প্রস্থ randomness. এইবার দ্বিতীয় পাত্র থেকে একটা যেকোনো বল তোলা হল, random-ভাবে। এই হল দ্বিতীয় প্রস্থ randomness. এই দ্বিতীয় প্রস্থ randomness-এর পরে দেখা গেছে যে একটা সাদা বল উঠেছে। Fig 17 দ্যাখো। আমাদের বার করতে হবে সবচেয়ে নীচের পথটার conditional probability.

এবার কিছু নাম দেওয়া যাক। প্রথম পাত্র থেকে দ্বিতীয় পাত্রে 5-টা বল পাঠানো হয়েছিল। এদের মধ্যে কিছু সাদা কিছু কালো। যদি কালোর সংখ্যা হয় j, তবে সাদার সংখ্যা হবে $5-j$. সেই অনুযায়ী--

> Let A_j be the event that j black and $5-j$ white balls are transferred from the first urn, for $j = 0, ..., 5$.
>
> Clearly, no other composition is possible. So A_j's are mutually exhaustive. Also no two compositions can occur simultaneously. So the A_j's are pairwise disjoint.

এরা হল প্রথম প্রস্থ randomness-এর পর কী কী হয়ে থাকতে পারে।

Let B be the event that the ball drawn from the second urn is white.

এটা হল দ্বিতীয় প্রস্থ randomness-এর পর কী হয়েছে।
এইবার probability-গুলো বার করব--

For each $j = 0, ..., 5,$ we have

$$P(A_j) = \frac{\binom{5}{j}\binom{5}{5-j}}{\binom{10}{5}} > 0.$$

[[Because:

- 5 balls may be chosen from 10 balls in $\binom{10}{5}$ways.
- j balls may be chosen from 5 black balls in $\binom{5}{j}$ways.
- $5-j$ balls may be chosen from 5 white balls in $\binom{5}{5-j}$ways.

]]

Then $P(B|A_j) = \frac{5-j}{5}$.

কারণ A_j ঘটেছে মানে দ্বিতীয় পাত্রের 5-টা বলের মধ্যে ঠিক $5-j$-টা হল সাদা।

So for different values of j we have the table:

j	0	1	2	3	4	5
$P(A_j)$	$1/\binom{10}{5}$	@$5/\binom{10}{5}$	$100/\binom{10}{5}$	$100/\binom{10}{5}$	$25/\binom{10}{5}$	$1/\binom{10}{5}$
$P(B\|A_j)$	1	$\frac{4}{5}$	$\frac{3}{5}$	$\frac{2}{5}$	$\frac{1}{5}$	0

Clearly $P(B) > 0$, since we had at least one white ball originally.

$\therefore$By Bayes' theorem,

$$\begin{aligned} P(A_0|B) &= \frac{P(A_0)P(B|A_0)}{\sum_0^5 P(A_j)P(B|A_j)} \\ &= \frac{1\times 5}{1\times 5 + 25\times 4 + 100\times 3 + 100\times 2 + 25\times 1 + 1\times 0} \\ &= \frac{1}{126}, \end{aligned}$$

which is the required answer.

■

Exercise 3: From an urn containing 3 white and 5 black balls, 4 balls are transferred into an empty urn. From this urn a ball is drawn and is found to be white. What is the probability that out of the 4 balls transferred 3 are white and 1 is black.? **(2007.1c)**

HINT:
আগের অংকের অনুকরণে কর দেখি। ■

Example 15: There are three coins, identical in appearance, one of which is unbiased and the other two are biased with probabilities $\frac{1}{3}$ and $\frac{2}{3}$, respectively, for a head. One coin is taken at random and tossed twice. If a head appears both times, what is the probability that the unbiased coin was chosen?[5] **(2010.1e)**

SOLUTION:

> Let A_1 be the event that the unbiased coin is selected.
> Let A_2 be the event that the coin with $P(H) = \frac{1}{3}$ is selected.
> Let A_3 be the event that the coin with $P(H) = \frac{2}{3}$ is selected.
>
> Let B be the event that the tosses produce the outcome HH.
>
> Then $P(A_1) = P(A_2) = P(A_3) = \frac{1}{3}$.
>
> Also $P(B|A_1) = \left(\frac{1}{2}\right)^2$, $P(B|A_2) = \left(\frac{1}{3}\right)^2$, and $P(B|A_3) = \left(\frac{2}{3}\right)^2$.

আমাদের বার করতে বলেছে $P(A_1|B)$.

> So, by Bayes' theorem,
>
> $$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)} = \cdots$$
>
> মাঝের ধাপ কয়টা তোমার করার জন্য রেখে দিলাম।
>
> $$= \frac{1}{9},$$
>
> which is the required answer.

■

Answers

3. $\frac{1}{7}$.

Chapter IV
Moments

DAY 12
Expectation

12.1 সহজভাবে বোঝা

ক্যাসিনোতে একটা জনপ্রিয় খেলা হল roulette (রুলেট)। সিনেমা-টিনেমাতেও যখনই ক্যাসিনো দেখায় তখনই এই খেলাটা সাধারণতঃ অন্ততঃ একবার দেখাবেই। এর প্রধান উপকরণ হল একটা রংচঙে চাকা, আর তার লাগোয়া একটা রংচঙে টেবিল। Fig 1 দ্যাখো। চাকাটার পরিধির গায় খোপ খোপ করা থাকে, আর খোপের পাশে পাশে সংখ্যা লেখা $1, 2, ..., 36$. এদের মধ্যে কিছু লাল, আর কিছু কালো। আর দুটো সংখ্যা থাকে সবুজ রঙের, 0 আর 00. টেবিলটার উপরে এই সংখ্যাগুলোরই নানারকম subset লেখা থাকে। Fig 2-এ বড়ো করে দেখিয়েছি। যেমন, একটা subset হয়তো সব লাল সংখ্যাগুলোর (ছবিতে ছাই ছাই রং মানে লাল), আরেকটা হয়তো সব বিজোড় সংখ্যাগুলোর, এইরকম। একটা subset-এর নাম হল "1st dozen" অর্থাৎ $\{1, 2, ..., 12\}$. তিনটে আছে "2 to 1". এরা হল টেবিলের একেকটা সারির জন্য, যেমন সবচেয়ে নীচের সারিটা হল $\{1, 4, 7, ..., 34\}$. এই খেলাটা খেলার জন্য টিকিট কাটতে হয়, একজন যতগুলো খুশি কাটতে পারে। ধরো তুমি একটা টিকিট নিলে। সেই টিকিটটা তোমাকে রাখতে হবে টেবিলে আঁকা কোনো একটা subset-এর উপর, মানে তুমি ওই subset-টার উপর বাজি ধরছ। যেমন ধরো Fig 3-এ বাজি ধরা হয়েছে যাবতীয় লাল রঙের সংখ্যার উপর। এবার ক্যাসিনোর লোক এসে চাকাটার উপর একটা বল রাখবে, তারপর চাকাটাকে বাঁইবাঁই করে ঘুরিয়ে দেবে। সাথে সাথে বলটাও ধাক্কা খেতে খেতে ঘুরবে। এক সময়ে চাকাটা থামলে বলটাও টুপ করে কোনো একটা খোপে পড়ে যাবে। যদি সেই খোপের সংখ্যাটা তোমার subset-এ থাকে তবে তুমি বাজি জিতবে এই নিয়মে--

$$\frac{36}{\text{তোমার subset-এর সাইজ}} \times \text{কত টাকার টিকিট কেটেছিলে}$$

যদি বলটা তোমার subset-এ না পড়ে তবে তুমি কিছুই পাবে না, মানে টিকিটের টাকাটা গচ্চা! চাইলে তুমি অনেক টিকিট কেটে একেকটাকে একেক জায়গায় রাখতে পারো। টিকিটও আবার পাওয়া যায় বিভিন্ন দামের। বুঝতেই পারছ যে, কোন দামের কটা টিকিট কেটে কোথায় কোথায় রাখলে ভালো হবে সেটা হিসেব করতে গিয়ে মাথা গুলিয়ে যাওয়া স্বাভাবিক। তার উপর আবার ক্যাসিনোর অন্দরমহলে হাস্যলাস্যের এত তরল তুফান চলে যে, তাতে মাথা ঠাণ্ডা হবার বড় একটা সুযোগ পায় না। এবার মনে কর তোমাকে একজন এই উপদেশ দিল--

দুটো দশ টাকার টিকিট কাটো। একটা দিয়ে বাজি ধরো জোড় সংখ্যার উপর, অন্যটাতে লাল সংখ্যাগুলোর উপরে।

Fig 1



Fig 2

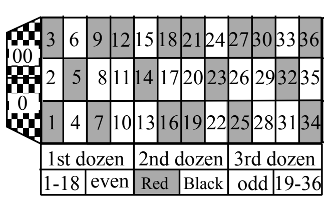
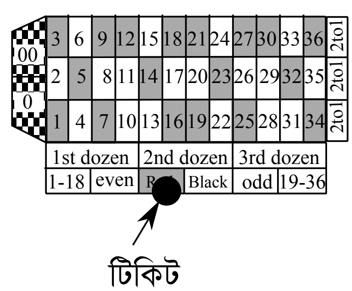


Fig 3



মাথা ঠাণ্ডা করে ভেবে বলো তো সেটা কি বুদ্ধিমানের কাজ হবে? এরকম ক্ষেত্রে সমস্যা এই যে, তোমার পকেট থেকে যেটা বেরোচ্ছে (মানে $2 \times 10 = 20$ টাকা) সেটা জানা থাকলেও ঠিক কতটা ঢুকবে সেটা জানা যাচ্ছে না। যেমন--

- যদি বলটা গিয়ে এমন কোনো সংখ্যায় পড়ে যেটা লাল এবং even (যেমন 16), তবে তুমি দুটো বাজিতেই জিতবে। প্রথম বাজিতে জিতলে পাবে $\frac{36}{20} \times 10 = 18$ টাকা, কারণ 0 আর 00-কে নিয়ে মোট even-এর সংখ্যা হল 20. দ্বিতীয় বাজি থেকে পাবে $\frac{36}{18} \times 10 = 20$ টাকা, কারণ লালের সংখ্যা হল 18. তার মানে তুমি পাচ্ছ মোট $20 + 18 = 38$ টাকা, এর থেকে টিকিট দুটোর দাম বাদ দিলে থাকে $38 - 2 \times 10 = 18$ টাকা, যেটা তোমার নগদ লাভ।

- কিন্তু যদি 15 পড়ে, তবে দুটোতেই হার, মানে টিকিটের 20 টাকাই গচ্চা!

- আবার যদি 7 পড়ে তবে খালি দ্বিতীয় বাজিটায় জিত, মানে পাচ্ছ খালি 20 টাকা। সেক্ষেত্রে তোমার লাভ লোকসান কাটাকাটি হয়ে কিছুই থাকছে না।

- আবার যদি 24 পড়ে তবে খালি প্রথম বাজিটায় জিত, মানে পাচ্ছ খালি 18 টাকা। সেক্ষেত্রে তোমার লাভ দাঁড়াচ্ছে $18 - 2 \times 10 = -2$ টাকা, অর্থাৎ পকেট থেকে 2 টাকা লোকসান।

বুঝতেই পারছ যে, এই চার ধরণের জিনিসের বাইরের আর কিছু ঘটতে পারে না। এই হিসেবগুলো সহজ, সমস্যা হল কোনটা হতে যাচ্ছে সেটা না জেনে সিদ্ধান্ত নেওয়া। এরকম অবস্থায় কীভাবে এগোতে হয়, তা তো আমরা দ্বিতীয় অধ্যায় থেকেই শিখেছি। এখানে খেলাটা হল একটা random experiment. প্রথম কাজ হল তার যাবতীয় outcome-এর একটা তালিকা বানিয়ে নেওয়া, যাকে sample space নাম দিয়েছিলাম। এখানে sample space-টা হল

$$\{0, 00, 1, ..., 36\}.$$

এর প্রত্যেকটা ক্ষেত্রে তুমি বাজিদুটো থেকে মোট কত পাচ্ছ, সেটা হল একটা random variable. যদি outcome-টাকে বলি ω তবে তার জন্য তোমার লাভের পরিমাণকে নাম দেওয়া যাক $X(\omega)$. এটা হল তোমার মোট লাভ, টিকিটের দামটা বিয়োগ করার পর। বিভিন্ন ω-র জন্য $X(\omega)$-র একটা টেবিল বানিয়ে নিলে সুবিধা হবে। টেবিলটা একটু লম্বা হবে, বইয়ের পাতায় পুরোটা একবারে আঁটবে না, তাই তিনটে অংশে ভেঙে দেখিয়েছি--
টেবিলের প্রথম অংশ--

সংখ্যা (ω)	-2	00	1	2	3	4	5	6	7	8	9	10	11	12
Even?	হ্যাঁ	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ
লাল?	না	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না	না	হ্যাঁ
লাভ ($X(\omega)$)	-2	-2	0	-2	0	-2	0	-2	0	-2	0	-2	-20	18

দ্বিতীয় অংশ--

সংখ্যা (ω)	13	14	15	16	17	18	19	20	21	22	23	24
Even?	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ
লাল?	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না
লাভ ($X(\omega)$)	-20	18	-20	18	-20	18	0	-2	0	-2	0	-2

তৃতীয় অংশ--

সংখ্যা (ω)	25	26	27	28	29	30	31	32	33	34	35	36
Even?	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ
লাল?	হ্যাঁ	না	হ্যাঁ	না	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ	না	হ্যাঁ
লাভ ($X(\omega)$)	0	-2	0	-2	-20	18	-20	18	-20	18	-20	18

টেবিলের প্রথম লাইনে লিখে নিয়েছি বলটা কোথায় কোথায় পড়তে পারে। তারপরের দুটো লাইনে হিসেবে রেখেছি দুটো বাজির কোনটায় কীরকম হারজিত হচ্ছে। যদি even হয় তবে প্রথম বাজিতে জিত, লাল হলে দ্বিতীয় বাজিতে জিত। আমরা এক্ষুণি যে হিসেব করলাম তা থেকেই জানি যে, দুটো বাজিতেই জিতলে লাভ হবে 18 টাকা, খালি প্রথমটায় জিতলে লাভ হবে 0 টাকা, খালি দ্বিতীয়টায় জিতলে -2 টাকা, আর দুটোতেই হারলে -20 টাকা। সেইগুলোই লিখেছি টেবিলের শেষ লাইনে।
এইবার কল্পনা কর যেন তুমি খেলাটা বার বার খেলবে। যদি চাকাটায় কোনো কারচুপি না থেকে থাকে তবে ধরা যায় যে, 38-খানা সংখ্যার মধ্যে প্রতিটাই মোটামুটি সমান হারে (অর্থাৎ $\frac{1}{38}$ হারে) আসবে। তাহলে প্রতিক্ষেত্রে তোমার লাভটাকে $\frac{1}{38}$ দিয়ে গুণ করে যোগ করে দিলে একটা ধারণা পাবে যে অনেকবার খেললে প্রতি খেলায় গড়ে তুমি কতটা লাভ করবে। সবগুলো ধরে ধরে যোগ করতে গেলে ব্যাপারটা বেশ লম্বা দাঁড়াবে--

$$\begin{aligned}&\tfrac{1}{38}(-2+-2+0-2+0-2+0-2+0-2+0-2-20+18+\\&-20+18-20+18-20+18+0-2+0-2+0-2\\&+0-2+0-2-20+18-20+18-20+18-20+18).\end{aligned}$$

বাপ্ রে, কি লম্বা জিনিস! তবে লক্ষ করো যে, যতই লম্বা হোক, আছে তো মোটে চারটে সংখ্যা, $-2, 0, 18$ আর $-20.$ সেগুলোই বারবার করে ঘুরে ফিরে এসেছে। সুতরাং একই সংখ্যাদের একসঙ্গে লিখলে ব্যাপারটা অনেক সহজে হবে। গুণে দ্যাখো, -2 আছে 10-টা, 0 আছে 12-টা, এবং 18 আর -20 দুজনেই আছে ঠিক 8 বার করে। সুতরাং পুরো জিনিসটা দাঁড়াচ্ছে--

$$\tfrac{1}{38}(12\times(-2)+10\times 0+8\times 18+8\times 20)\approx -1.05,$$

যেটা < 0, মানে প্রতি খেলায় গড়ে তোমার পকেট থেকে একটাকা পাঁচ পয়সা বেরিয়ে যাচ্ছে! সুতরাং খেলাটা তোমার পকেটের পক্ষে ভালো নয় মোটেই।
এইখানে তুমি একটা আপত্তি তুলতে পারো এই বলে যে, আমি তো মোটে একবার খেলব কিনা ভাবছিলাম। এর মধ্যে খামোখা বারবার খেলার প্রশ্ন এনে ফেললাম কেন? এমন তো হতেই পারে যে একবার খেলেই জিতে গেলাম, তখন বাজির টাকাটা পকেটে পুরে ঘরের ছেলে[1] ঘরে ফিরে গেলেই হল! ঠিক কথা। যে কোনো random experiment-ই যদি খালি একবার (বা অল্প কয়েকবার) করা হয়, তার outcome সম্পর্কে কিছুই জোর দিয়ে বলা যায় না। যদি যেত তবে সেটা আর random experiment থাকত না। কিন্তু এ বইয়ের প্রথম অধ্যায়েই বলেছি যে random experiment-টা বারবার করলে ক্রমশঃ তার ভিতর statistical regularity ফুটে ওঠে। আমাদের কায়দাটা সেই statistical regularity-র ভিত্তিতেই কাজ করে। অল্প কয়েকবার করলে কোনো প্যাটার্ণ বোঝা যায় না। কিন্তু তাও তুমি অসহিষ্ণু হয়ে বলতে পারো যে, খামোখা statistical regularity আনার জন্য কে আর ওই একই খেলা বার বার খেলতে যাচ্ছে! সুতরাং এই অংকটা করে বাস্তবে কার উপকার? উত্তরটা খুবই সহজ--উপকারটা ক্যাসিনোর মালিকের। তুমি না হয় একবার খেলেই বেরিয়ে এলে, কিন্তু এরকম বহু খদ্দের যদি প্রত্যেকেই একবার খেলেও বেরিয়ে যায়, তাও তো খেলাটা বহু বহুবার খেলা হয়ে যায়, সুতরাং প্রতিটা খেলা থেকেই গড়ে ক্যাসিনো মালিকের লাভ হয়। এর মধ্যে যে কয়জন খদ্দের লাভের টাকা পকেটে নিয়ে ফেরে, তারা ভাবে যে, সে টাকা বুঝি ক্যাসিনো মালিকের পকেট থেকে এল। মোটেই তা নয়। সে টাকা এসেছে অন্য কোনো হতভাগ্য খদ্দেরের পকেট থেকে, ক্যাসিনোর মালিক খালি লাভই করে চলেছে।

12.2 অংকের ভাষায় লেখা

এক্ষুণি যে লম্বা যোগটা লিখেছিলাম সেটার দিকে আরেকবার তাকানো যাক। আমরা প্রতিটা outcome নিচ্ছিলাম, মানে বলটা কোথায় কোথায় পড়তে পারে। এদের যে কোনো একটা ধরো ω. তার জন্য $X(\omega)$ বার করছিলাম। সেটাকে সেই outcome-এর probability, $P(\{\omega\})$, দিয়ে গুণ করে দিচ্ছিলাম। এখানে সব গুলো $P(\{\omega\})$-ই ছিল $\frac{1}{38}$. সেটা অন্য অংকে নাও হতে পারত (যেমন চাকাটা যদি সামান্য হেলে ঘোরে তবে একদিকের probability একটু বেড়ে যাবে)। তারপরে সবাইকে যোগ করে দিয়েছিলাম। এর ফলে যেটা পেয়েছিলাম সেটা হল "বহুবার একই খেলা খেললে মোটামুটি গড়ে কত লাভ প্রত্যাশিত"। ইংরাজিতে "প্রত্যাশা"-কে বলে expectation. তাই আমরা এই লম্বা যোগফলটাকে বলা হয় X-এর **expectation,** আর লেখা হয় $E(X)$, মানে

$$E(X)=\sum_\omega X(\omega)P(\{\omega\}).$$

"Expectation" না বলে অনেক সময়ে একে X-এর **mean** বা **average**-ও বলা হয়।

[1] বা ঘরের মেয়ে।

আমাদের অংকে এই sum-টায় 38-টা term ছিল। ফলে পুরো sum-টা বেজায় লম্বা হয়ে গিয়েছিল। তাই আমরা X-এর একই value-গুলোকে এক জায়গায় করে লিখে পুরো sum-টাকে সংক্ষিপ্ত করে নিয়েছিলাম। এর ফলে পেয়েছিলাম--

$$\tfrac{1}{38}(10\times(-2)+12\times 0+8\times 18+8\times 20)=\tfrac{10}{38}\times(-2)+\frac{12}{38}\times 0+\frac{8}{38}\times 18+\frac{8}{38}\times 20.$$

একটু চিন্তা করলেই বুঝবে যে এখানে X-এর প্রতিটা value-কে তার probability দিয়ে গুণ করা হচ্ছে, যেমন -2-এর সঙ্গে যে $\frac{10}{38}$-টা গুণ হয়েছে সেটা আসলে $P(X=-2)$. এইটা ভালো করে বুঝে নাও। সুতরাং আমরা এভাবেও লিখতে পারি--

$$E(X)=\sum_x xP(X=x),$$

যেখানে sum-টা হচ্ছে X-এর যাবতীয় value-র উপরে। এইটাই সাধারণতঃ ব্যবহার করা সহজ, তাই আমরা এই ফর্মুলাটাকেই $E(X)$-এর সংজ্ঞা হিসেবে নেব।

DEFINITION: Expectation (finite version)

Let X be a random variable taking only finitely many values $x_1,...,x_k$. Then Its expectation is defined as $E(X)=\sum_{i=1}^k x_iP(X=x_i)$.

যদি X-এর মোট value-র সংখ্যা countably infinite হয়, তবে আন্দাজ করতে পারছ যে, একটা infinite series আসবে। এবং তার convergence নিয়ে আমাদের মাথা ঘামাতে হবে। এইখানে একটা প্যাঁচ আছে, খালি convergence হলে চলবে না, absolute convergence চাই। সেটা নীচের সংজ্ঞাটা দেখলেই বুঝবে।

DEFINITION: Expectation (infinite series version)

Let X be a random variable taking values $x_1,x_2,...$ Then its expectation is defined as

$$E(X)=\sum_{i=1}^{\infty} x_iP(X=x_i),$$

provided the series converges absolutely.

আশা করি "absolute convergence" মানে কী সেটা ভুলে যাও নি। এর মানে হল $\sum_{i=1}^{\infty}|x_iP(X=x_i)|<\infty$. কিন্তু এদিকে তো আমরা $E(X)=\sum_{i=1}^{\infty}x_iP(X=x_i)$ নিচ্ছি, তবে ওই বাড়তি absolute value-টা নেবার দরকার হচ্ছে কেন? তার কারণটা লুকিয়ে আছে Riemann's rearrangement theorem-এর মধ্যে, যেটা তোমরা 2nd year-এ শিখেছিলে (এবং সম্ভবতঃ এখন ভুলে গেছ)। আমাদের বাংলায়-বোঝানো-ইংরাজি-বই সিরিজের Real Analysis (volume 2)-তে এই theorem-টা নিয়ে বিশদ আলোচনা আছে। যাই হোক, এখানে সংক্ষেপে মনে করিয়ে দিই--

> ধরো $\sum_{i=1}^{\infty}x_iP(X=x_i)$-টা converge করে, কিন্তু $\sum_{i=1}^{\infty}|x_iP(X=x_i)|=\infty$. সেক্ষেত্রে Riemann's rearrangement theorem বলছে যে, তুমি যদি x_i-দের ক্রমটাকে ওলটপালট করে দাও, তবে $\sum_{i=1}^{\infty}x_iP(X=x_i)$-এর value অন্য কিছু হয়ে যেতে পারে।

কিন্তু বুঝতেই পারছ যে, গড় বার করার সময়ে কাকে আগে কাকে পরে লেখা হচ্ছে সেটা গুরুত্বপূর্ণ হওয়া উচিত নয়। কিন্তু যদি absolutely convergent হয়, তবে সে সমস্যা থাকে না। সেই জন্যই আমরা absolute convergence শর্তটা চাপিয়েছি। এই প্রসঙ্গে দুটো উদাহরণ দিই, যেখানে $E(X)$-টা exist করে না। উদাহরণ দুটো বোঝার জন্য তোমাকে তিনটে infinite series-এর কথা মনে করতে হবে--

- $\sum_1^\infty \frac{1}{n}$, যেটা ∞-তে diverge করে,
- $\sum_1^\infty \frac{1}{n^2}$ যেটা converge করে, এর limit-টার নাম দিলাম ধরো c.
- $\sum_1^\infty \frac{(-1)^n}{n}$, যেটা converge করে।

Example 1: ধরো আমরা একটা random variable নিলাম X, যেটা $1, 2, 3, \ldots$ ইত্যাদি value নিতে পারে। এর PMF-টা নিলাম ধরো--

$$P(X = n) = \begin{cases} \frac{1}{cn^2} & \text{if } n \in \mathbb{N} \\ ) & \text{otherwise.} \end{cases}$$

এটা যে সত্যিই একটা PMF তা নিয়ে সন্দেহ নেই, কারণ প্রতিটা probability-ই positive,এবং যোগ করলে $\sum_1^\infty \frac{1}{cn^2} = \frac{1}{c}\sum_1^\infty \frac{1}{n^2} = 1$. এবার দেখি $E(X)$ আছে কিনা। এর জন্য এই infinite series-টার absolute convergence দেখতে হবে--

$$\sum_1^\infty nP(X = n) = \sum_1^\infty n\frac{1}{cn^2} = \frac{1}{c}\sum_1^\infty \frac{1}{n} = \infty.$$

যাঃ, diverge করে গেল! সুতরাং $E(X)$-টা exist করছে না। ■

Example 2: এবার একটা random variable দেখি, Y, যেটা এই value-গুলো নিতে পারে--

$$-1, 2, -3, 4, -5, 6, \ldots$$

মানে সেই সব সংখ্যা যাদেরকে $(-1)^n n$ আকারে লেখা যায়, যেখানে $n \in \mathbb{N}$. এখানেও PMF-টা অনেকটা আগের উদাহরণের মতই নেব--

$$P(Y = n) = \frac{1}{cn^2}, \quad n \in \{-1, 2, -3, 4, \ldots\}.$$

এখানে $E(Y)$-এর existence পরীক্ষা করে দেখি। তার জন্য এই infinite series-টা নিয়ে কাজ করতে হবে--

$$\sum_1^\infty (-1)^n nP(Y = (-1)^n n) = \sum_1^\infty (-1)^n n\frac{1}{cn^2} = \frac{1}{c}\sum_1^\infty \frac{(-1)^n}{n}.$$

এই series-টা converge করে। কিন্তু আমাদের চাই absolute convergence. দুঃখের ব্যাপার যে, একবার absolute value নিলে পরে সেই $\sum \frac{1}{n}$-এই গিয়ে পড়ব, যেটা converge করে না। তাই এখানেও expectation-টা exist করছে না। ■

আমরা expectation-এর দু ধরণের সংজ্ঞা দিয়েছি--এক, যেখানে random variable-টা খালি finite-সংখ্যক value নিতে পারে। আর দুই, যেখানে countably infinite-সংখ্যক value নিতে পারে। এবার আলোচনা করব সেই সব random variable-এর কথা যাদের density আছে। এখানে outcome-এর সংখ্যা হবে uncountably infinite. তাই আর infinite series দিয়ে কাজ হবে না। এখানে আমরা integration লাগাব। সংজ্ঞাটা হবে এইরকম--

> DEFINITION: **Expectation (density version)**
>
> Let X be a random variable with density $f(x)$. If $\int_{-\infty}^{\infty} |xf(x)|dx < \infty$, then we define expectation of X as
>
> $$E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

12.2.1 Expectation of a function

যেভাবে আমরা $E(X)$ বার করলাম, সেই কায়দায় X-এর কোনো function-এর expectation-ও বার করা যায়। ধরো $h(x)$ কোনো একটা function. তবে $h(X)$ আবার একটা random variable হবে। যদি X একটা discrete random variable হয় যেটা $x_1, x_2, ...$ ইত্যাদি value নেয়, এবং $p_i = P(X = x_i)$ হয় $(i = 1, 2, ...)$, তবে

$$E(h(X)) = \sum_i h(x_i)p_i$$

হবে, যদি অবশ্য $\sum_i |h(x_i)p_i| < \infty$ হয় (না হলে বলব $E(h(X))$-টা exist করে না)। একইভাবে, যদি X-এর density থাকে $f(x)$, তবে

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)f(x)dx,$$

যদি $\int_{-\infty}^{\infty} |h(x)f(x)|dx < \infty$ হয়।
এবার হাতেকলমে একটা অংক করা যাক।

Example 3: A box contains a white and b black balls. If c balls are drawn, then find the expectation of the number of white balls drawn.[4] **(2005.3b)**
Solution:

> Let X denote the number of white balls in the sample.
> Then
> $$P(X = x) = \frac{\binom{a}{x}\binom{b}{c-x}}{\binom{a+b}{c}} \text{ for } x = 0, 1, ..., c.$$

এইটা কী করে এল বুঝে নিই। বাক্সে মোট বলের সংখ্যা ছিল $a + b$. তার থেকে c-টা নেওয়া যায় $\binom{a+b}{c}$-ভাবে (এরই আরেক নাম ${}^{a+b}C_c$)। এর মধ্যে আমরা খালি সেইসব কেসগুলো নিয়েই মাথা ঘামাচ্ছি যখন ঠিক x-টা সাদা বল, আর ঠিক $c - x$-টা কালো বল আছে। যেহেতু a-খানা সাদা বল আছে, তার থেকে x-টা নেওয়া যায় $\binom{a}{x}$-ভাবে, আর b-খানা কালো বল থেকে $c - x$-টা নেওয়া যায় $\binom{b}{c-x}$-ভাবে।

> Some of these probabilities may be 0.

So

$$E(X) = \sum_{x=0}^{c} xP(X=x)$$

$$= \sum_{x=0}^{c} \frac{x\binom{a}{x}\binom{b}{c-x}}{\binom{a+b}{c}}$$

নীচের ওই $\binom{a+b}{c}$-টায় কোনো x নেই, ওটাকে sum-এর বাইরে নিয়ে আসি--

$$= \frac{1}{\binom{a+b}{c}} \sum_{x=0}^{c} x\binom{a}{x}\binom{b}{c-x}$$

এবার একটা কৌশল করব--

আমরা জানি যে, $r\binom{n}{r} = n\binom{n-1}{r-1}$ হয় (সত্যিই জানো তো?)

কিন্তু এর জন্য $r \geq 1$ হতে হয়, $r = 0$ হলে ওই নীচের $r-1$-টা < 0 হয়ে যাবে।

আমাদের এখানে $x = 0$ থেকে sum-টা শুরু, কিন্তু আসলে $x = 0$-র term-টা শূন্য।

তাই ওটাকে sum থেকে বাদ দেওয়া যায়--

$$= \frac{1}{\binom{a+b}{c}} \sum_{x=1}^{c} x\binom{a}{x}\binom{b}{c-x}$$

$$= \frac{1}{\binom{a+b}{c}} \sum_{x=1}^{c} a\binom{a-1}{x-1}\binom{b}{c-x} \quad \left[\because r\binom{n}{r} = n\binom{n-1}{r-1} \text{ for } r \geq 1.\right]$$

$$= \frac{a}{\binom{a+b}{c}} \sum_{x=1}^{c} \binom{a-1}{x-1}\binom{b}{c-x}$$

$$= \frac{a}{\binom{a+b}{c}} \sum_{y=0}^{c-1} \binom{a-1}{y}\binom{b}{c-y-1} \quad \left[\text{putting } y = x-1\right]$$

$$= \frac{a\binom{a+b-1}{c-1}}{\binom{a+b}{c}} \quad [\text{by binomial theorem identity}]$$

এখানে আমরা এই জিনিসটা লাগালাম--

যদি $m, n \in \mathbb{N}$ হয় তবে $\binom{m}{0}\binom{n}{r} + \binom{m}{1}\binom{n}{r-1} + \binom{m}{2}\binom{n}{r-2} + \cdots$ করে গেলে $\binom{m+n}{r}$ হয়।

$$= \frac{ac}{a+b},$$

which is the required answer.

■

12.3 Properties

আমরা expectation-এর তিন রকম সংজ্ঞা দিয়েছিলাম--এক, finite sum দিয়ে, দুই, infinite series দিয়ে, আর তিন, integral দিয়ে। এই তিনক্ষেত্রেই কয়েকটা ধর্ম পালিত হয়, যেগুলো এবার বলব। বোঝার সুবিধার জন্য roulette-এর

উদাহরণটা মাথায় রাখলে সুবিধা হবে। সেখানে আমরা মোট লাভ, X, বার করেছিলাম এইভাবে--

$$X = Y + Z - T,$$

যেখানে Y হল প্রথম বাজি থেকে পাওয়া টাকা, Z হল দ্বিতীয় বাজির টাকা, আর T হল টিকিটের খরচা। এখানে X-এর মত Y, Z, T-ও হল random variable. অবশ্য তুমি আপত্তি তুলতে পারো যে, T তো সব সময়েই একই থাকবে, ওর মধ্যে random-এর কী আছে? ঠিক কথা। কিন্তু ওটাকেও random variable-এর একটা বিশেষ কেস হিসেবে ধরা হয়, এটা এমন একটা random variable যেটা আসলে constant. কথাটা "সোনার পাথরবাটি"-র মত শোনাতে পারে, variable আবার constant হয় কী করে? কিন্তু আসলে এতে কোনো সমস্যা নেই, variable মানে যেটা পরিবর্তিত হতে পারে, পরিবর্তিত হবেই এমন কিন্তু কোনো কথা নেই! এরকম জিনিস তুমি আগেও দেখেছ-- constant function, যেমন $f(x) \equiv 5$. এটাকে x-এর function বলছি বটে, কিন্তু x যাই value নিক না কেন এটা সবসময়েই 5হয়ে বসে থাকবে। একইভাবে এখানে $T(\omega) \equiv 20$, চাকায় যে সংখ্যাই আসুক না কেন, দুটো টিকিটের দাম সেই 20 টাকাই থাকবে।
এইবার তবে ধর্মগুলো একে একে বলি--

1. যদি random variable-টা আসলে একটা constant হয়, ধরো c, তবে তার expectation-ও হবে সেই c-ই, মানে $E(c) = c$. যেমন আমাদের উদাহরণে $E(T) = 20$.

2. যদি $E(X)$-টা exist করে, আর a কোনো সংখ্যা হয়, তবে $E(aX)$এবং $E(X + a)$-ও exist করবে, এবং $E(aX) = aE(X)$ আর $E(X + a) = E(X) + a$ হবে। যেমন যদি লাভের হিসেবটা ভারতীয় টাকায় না করে নেপালী টাকায় করি, যেখানে ভারতের 1 টাকা মানে নেপালের a টাকা, তবে expectation-টাও a দিয়ে গুণ হয়ে যাবে।

3. ধরো দুটো random variable আছে Y আর Z, যাদের জন্য random experiment-টা একই। যেমন roulette-এর উদাহরণের Y, Z. এখানে Y, Z দুজনেই একই random experiment মানে roulette থেকে পাওয়া। সেক্ষেত্রে যদি $E(Y), E(Z)$ দুজনেই exist করে তবে $E(Y + Z)$এবং $E(Y - Z)$-ও exist করবে, এবং $E(Y + Z) = E(Y) + E(Z)$ আর $E(Y - Z) = E(Y) - E(Z)$ হবে।

এই ধর্ম কয়টা তিনটে সংজ্ঞার ক্ষেত্রেই প্রমাণ করা অত্যন্ত সহজ। চেষ্টা করেই দ্যাখো! তবে শেষেরটার ব্যাপারে একটু সাবধান করে দিই--

- $E(YZ)$ কিন্তু $E(Y)E(Z)$-এর সমান নাও হতে পারে! আমাদের roulette-র উদাহরণেই এরা সমান হবে না।
- যদি $E(Y)$ আর $E(Z)$ কেউই exist না করে তাও কিন্তু $E(Y + Z)$ বা $E(Y - Z)$ দিব্যি exist করতে পারে। এরকম উদাহরণ বানানো খুব সহজ। একটু ধরিয়ে দিই--একটু আগে আমরা দুটো random variable দেখিয়েছিলাম যাদের expectation-টা exist করে না, তাদের যে কোনো একটাকে Y নাও, আর এবার ভাবো তো কী করে Z নেবে! উত্তরটা ফুটনোটে[2] দিলাম। উত্তরটা বোকা-বোকা রকমের সহজ, দেখে নিলে নিজেরই লজ্জা করবে কিন্তু!

এবার আরেকটা ধর্মের কথা বলব যার নাম হল monotonicity. যদি X, Y দুটো random variable হয়, তবে $X \geq Y$ হলে $E(X) \geq E(Y)$ হতে বাধ্য। এটা অবশ্য নিতান্তই আল্‌গাভাবে বলা হল, শর্ত-টর্ত বাদ দিয়ে। সব কিছু গুছিয়ে লিখলে হবে--

Monotonicity of expectation

Let X, Y be two random variables, both defined on the same random experiment with sample space S. Let $E(X), E(Y)$ both exist. If $\forall \omega \in S \quad X(\omega) geq Y(\omega)$, then $E(X) \geq E(Y)$.

[2] $Z = Y$ নাও, তবে $Y - Z = 0$, যার expectation-ও 0.

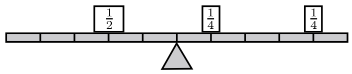


Fig 4

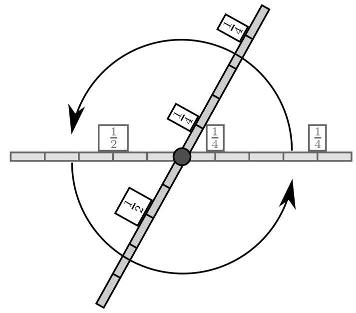


Fig 5

12.4 ছবি দিয়ে ভাবা

এবার আমরা একটা উপমার কথা শিখব যেটা আমাদের ছবি দিয়ে ভাবতে সাহায্য করবে। মনে করো একটা random variable আছে X, যেটা খালি তিনটে value নিতে পারে, $-2, 1$ আর 4. এদের probability ধরো যথাক্রমে $\frac{1}{2}, \frac{1}{4}$ এবং $\frac{1}{4}$. তাহলে $E(X)$ বার করা সহজ--

$$E(X) = -2 \times \tfrac{1}{2} + 1 \times \tfrac{1}{4} + 4 \times \tfrac{1}{4} = \tfrac{1}{4}.$$

এটাকে physics দিয়ে ভাবা যায়--মনে করো একটা স্কেল আছে, যার গায় $-2, 1$ আর 4-এ দাগ কাটা আছে। প্রত্যেকটা দাগে একটা করে ওজন বসানো আছে, ওজনগুলো probability-র সমান (Fig 4)। এইবার যদি এই ওজনসুদ্ধ স্কেলটাকে 0 দাগে ব্যালান্স করে বসাতে যাও তবে কী হবে? ব্যালান্স করে থাকবে? নাকি বাঁদিকে কাত হয়ে যাবে? নাকি ডানদিকে? সেটা স্থির করার কায়দা আমরা হায়ার সেকণ্ডারীতেই শিখি। প্রত্যেকটা ওজনের moment বার করতে হয়। ফর্মুলাটা হল ওজন $\times$ দূরত্ব। এখানে দূরত্বর সঙ্গে একটা চিহ্ন জড়িত আছে, 0-র বাঁদিকে হলে negative, আর ডানদিকে হলে positive. মোট momentযদি > 0 হয় তবে স্কেলটা ঘড়ির কাঁটার মত হেলে যাবে, যদি < 0 তবে উল্টো দিকে হেলে যাবে। আর যদি ঠিক 0 হয়, তবে ব্যালান্স করে থাকবে। মোট moment মানে সবগুলো ওজনের moment-এর যোগফল। লক্ষ কর সেটাই ঠিক $E(X)$. সেই কারণে $E(X)$-এর আরেকটা নাম হল 1st moment around 0. এখানে "around 0" কেন বলছে বুঝতেই পারছ, যেহেতু আমরা 0-তে ব্যালান্স করা চেষ্টা করছি। আর "1st" বলার কারণ হল আমরা ওজনের সঙ্গে দূরত্ব গুণ করছি। যদি দূরত্বের square গুণ করতাম তবে সেটা হত 2nd moment around 0. হঠাৎ মনে হতে পারে খামোখা দূরত্বের square গুণ করতে যাব কেন? সেটাও physics-এর উপমা দিয়ে বোঝা যায়। সেই প্রসঙ্গে এক্ষুণি আসছি, কিন্তু তার আগে expectation-এর বিষয়ে একটা গুরুত্বপূর্ণ কথা সেরে নিই।

Example 4: ধরো X একটা random variable, যার expectation-টা exist করে। সুতরাং $E(X)$ একটা সংখ্যা। খবরদার, X যতই random হোক, $E(X)$ কিন্তু একটা সংখ্যা মাত্র (যদি exist করে)। ধরো $E(X) = a$. তবে $E(X - a)$ কত হবে?

SOLUTION:

$$E(X - a) = E(X) - a = a - a = 0.$$

■

ছবি দিয়ে ভাবলে এর ব্যাখ্যা হল--যদি তুমি স্কেলটাকে এক আঙুলে ব্যালান্স করে রাখতে চাও, তবে আঙুলটা রাখতে হবে $E(X)$-এ। এই একটা বিন্দুতে যেন ওজন-সুদ্ধ স্কেলটার যাবতীয় mass (ভর) কেন্দ্রীভূত হয়ে আছে, তাই এই বিন্দুটাকে বলে centre of mass, (ভরকেন্দ্র)। সেই অর্থে $E(X)$-এর আরেকটা ব্যাখ্যা হল--X-এর probability distribution-টাকে স্কেলের ওপর ওজন হিসেবে ভাবলে, $E(X)$ হল তার centre of mass.

Example 5: ধরো ওজনগুলো স্কেলটার গায় স্ক্রু দিয়ে আটকানো আছে, যাতে ওল্টালেও পড়ে না যায়। এবার তোমাকে বলা হল স্কেলটাকে চাকার মত ঘোরাতে, 0-কে কেন্দ্র করে। Fig 5 দ্যাখো। এর জন্য কত energy (মানে শক্তি) লাগবে?

SOLUTION: আন্দাজ করতে পারছ যে, ওজনগুলো 0 থেকে যত দূরে থাকবে কাজটা ততই কঠিন হবে, মানে বেশী energy লাগবে। কতটা বেশী, সেটা এইভাবে হিসেব করা যায়। ধরো তুমি ঘোরাচ্ছ w বেগে, এটা হল যাকে বলে angular velocity (কৌণিক গতিবেগ)। মানে সেকেণ্ডে কত ডিগ্রী ঘুরছে তার পরিমাপ। সবগুলো ওজনই স্কেলের সঙ্গে স্ক্রু দিয়ে আঁটা আছে বলে w সবার ক্ষেত্রেই একই। কিন্তু যে ওজনটা বেশী দূরে আছে, সেটা একইসময়ে বেশী দূরত্ব অতিক্রম করছে, ফলে তার velocity হবে বেশী। ফর্মুলাটা হল $v = wr$, যেখানে r হল দূরত্ব। আমরা জানি যে mass যদি m হয়, আর velocity হয় v, তবে তার kinetic energy হয় $\frac{1}{2}mv^2$, মানে আমাদের ক্ষেত্রে $\frac{1}{2}mr^2w^2$. লক্ষ কর কীভাবে m-এর সঙ্গে দূরত্বের square, মানে r^2, গুণ হয়ে গেল। Physics-এ সবগুলো ওজনের mr^2-এর যোগফলকে বলে moment of inertia. আমরা statistics-এ বলি **2nd moment around 0.** ■

এবার কয়েকটা নাম শিখে রাখি--

- যেকোনো $k = 1, 2, 3, ...$ ইত্যাদির জন্যই **k-th moment around 0**-র সংজ্ঞা হল $E(X^k)$.
- এই প্রসঙ্গে বলে রাখি যে moment around 0 না বলে আমরা অনেক সময়ে **raw moment**-ও বলি।
- Moment চাইলে আমরা যেকোনো সংখ্যাকে কেন্দ্র করেই নিতে পারি। যদি a কোনো একটা সংখ্যা হয়, তবে **k-th moment around a** হবে $E((X-a)^k)$.
- যদি a-র জায়গায় $E(X)$ নিই, তবে যে moment-গুলো পাব তাদের বলে **central moment.** অর্থাৎ k-th central moment-টা হল $E((X-E(X))^k)$.
- এদের মধ্যে $k = 2$ নিলে পাবো 2nd central moment, যার প্রচলিত নাম হল **variance.** এর মাহাত্ম্যের পরিচয় শীঘ্রই পাবে।

Exercise 1: ধরো X এমন একটা random variable যার k-th raw moment-টা exist করে। এই moment-টার নাম দিলাম μ_k. যদি $a \in \mathbb{R}$ হয়, তবে দেখাও যে, aX-এর k-th raw moment হবে $a^k\mu_k$. ■

Exercise 2: আগের অংকটা কি k-th central moment-এর বেলাতেও খাটবে? ■

এখানে একটা জিনিস বলে রাখি যেটা প্রমাণ করা শক্ত নয়--

> THEOREM
>
> If the k-th raw moment exists, then for every $\ell \leq k$, the ℓ-th raw moment and central moment must exist.

Exercise 3: ধরো X এমন একটা random variable যার k-th central moment-টা exist করে। এই moment-টার নাম দিলাম m_k. যদি $a \in \mathbb{R}$ হয়, তবে দেখাও যে, $X + a$-র k-th central moment-টা exist করে। সেটা বার করো তো। তোমার উত্তরে তুমি $m_1, m_2, ..., m_k$ ব্যবহার করতে পারো। ■

DAY 13
Central tendency, dispersion ইত্যাদি

13.1 Central tendency

একটা random variable কখন ঠিক কোন value নেবে বলা মুস্কিল। তাই আমরা অনেক সময়ে জানতে চাই যে "মোটামুটিভাবে" তার value-টা কত হবে। যেমন,

- কলকাতায় মে মাসে মোটামুটি কত তাপমাত্রা থাকে?
- JAM-এর জন্য একটা ভালো কোচিং ক্লাসে পড়তে কলকাতায় মাসে মোটামুটি কত খরচা পড়ে?

কিন্তু যদি জিজ্ঞাসা করি--

> একটা স্তন্যপায়ী প্রাণীর মোটামুটি কত ওজন হয়?

তবে সে প্রশ্নটার মানেই হয় না, কারণ স্তন্যপায়ী প্রাণী নানারকম হয়, এবং তাদের ওজনের মধ্যে আকাশপাতাল ফারাক। তিমি মাছও স্তন্যপায়ী, আবার নেংটি ইঁদুরটাও তাই। পৃথিবীর যাবতীয় স্তন্যপায়ীর ওজনের গড় বার করা যায় বটে, কিন্তু সেই সংখ্যাটার কোনো মানে হয় না।
যদি কোনো random variable-এর বেলায় তার একটা "মোটামুটি value" বলা যায়, তবে বলব যে সেই random variable-টার central tendency আছে। এখানে "মোটামুটি" কথাটা গুরুত্বপূর্ণ, কারণ এটা কোনো বিশেষ সংখ্যা নয়। এটা একটা ধারণা মাত্র। একটা পরিচিত উদাহরণ মাথায় রাখলে সুবিধা হবে। আমরা অনেক সময়েই দুজন ছাত্রের মধ্যে তুলনা করার সময়ে "কে বেশী মেধাবী" সেটা বিচার করি। এখানে "মেধা"-টা একটা ধারণা মাত্র। এক কথায় এর কোনো সংজ্ঞা দেওয়া কঠিন। তুমি পরীক্ষার নম্বর দিয়ে কে কত মেধাবী সেটা মাপার চেষ্টা করতে পারো। কিন্তু তাতে পুরো চিত্রটা পাবে না। পরীক্ষার সংকীর্ণ গণ্ডীর বাইরেও কে কতটা চিন্তাভাবনা করে সেটার হিসেব বাদ চলে যাবে। আবার কে কটা বই পড়েছে, সেটা দিয়েও মেধা মাপার চেষ্টা করা যেতে পারে। তাতে আবার কে কতটা সত্যিই বুঝছে, সেটার হিসেব পাওয়া যাবে না। অথচ এত সমস্যা সত্ত্বেও "মেধা" নামক ধারণাটা বুঝতে বা সেটা ব্যবহার করে তুলনামূলক আলোচনা করতে আমাদের অসুবিধা হয় না। Central tendency ব্যাপারটাও সেইরকম। অংক কষে central tendency মাপার জন্য নানা কায়দা ভাবা হয়েছে, কোনোটাই সব ক্ষেত্রে সম্পূর্ণ চিত্রটা দেয় না, কিন্তু তাও central tendency নিয়ে কাজ করতে অসুবিধা হয় না।
এবার তবে central tendency মাপার কয়েকটা কায়দা আলোচনা করা যাক।

13.1.1 Central tendency মাপার প্রথম কায়দা

যদি X একটা random variable হয়, যার central tendency মাপতে চাই, তবে সবচেয়ে জনপ্রিয় কায়দা হল $E(X)$ বার করা। ভুলে যেও না যে, central tendency আছে কি নেই সেটা নির্ভর করে কী নিয়ে কাজ করছি তার উপর। অন্ধের মত $E(X)$-এর ফর্মুলা লাগিয়ে দিলে যে উত্তর পাবে সেটা কিন্তু অর্থপূর্ণ নাও হতে পারে, যেমন পৃথিবীর যাবতীয় স্তন্যপায়ীদের ওজনের গড় বার করার মত। আবার যদি দ্যাখো যে $E(X)$-টা exist করছে না, তার মানেই এই নয় যে central tendency নেই। সেক্ষেত্রে central tendency মাপার অন্য কায়দা ব্যবহার করতে হবে, এই যা। সেরকম বিকল্প কায়দা আমরা শীঘ্রই দেখব।

13.1.2 Central tendency মাপার দ্বিতীয় কায়দা

এবার যে কায়দাটার কথা বলব সেটাও বস্তুতঃ $E(X)$-এরই রকমফের। $E(X)$ বার করার জন্য সাধারণতঃ আমাদের কাজ করতে হয় X-এর probability distribution নিয়ে। অনেক সময়ে সেটা সম্পূর্ণ জানা থাকে না। যেটা জানা থাকে সেটা হল X-এর কিছু value, ধরো $x_1, ..., x_n$. সেক্ষেত্রে আমরা central tendency মাপতে পারি

$$\bar{x} = \frac{1}{n}\sum_{1}^{n} x_i$$

দিয়ে। একে বলে **sample mean** বা **sample average.** এটাকে মনে করতে পারো যেন $E(X)$-এর statistical image.
এখানেও কিন্তু সেই একই সাবধানবাণী প্রযোজ্য-- একগুচ্ছ সংখ্যা পেলেই চোখ-কান বুঁজে তাদের average বার করে ফেললে সেইরকম average-এর কোনো অর্থ নাও দাঁড়াতে পারে। বস্তুতঃ এইরকম কাণ্ডের কথা ভেবেই সেই হাসির গল্পটা চালু

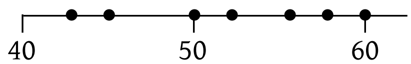


Fig 6

হয়েছে--statistician-রা নাকি এক পা গনগনে আগুনে আর অন্য পা বরফ জলে চুবিয়ে রেখে ভাবে যে, on an average বেশ আরামদায়ক তাপমাত্রা! ঠিক এই কাজটাই অবশ্য কেউ করে না, কিন্তু কখনোসখনো এমন আলোচনা শোনা যায় যেখানে কোনো "পণ্ডিত" ব্যক্তি টাটা বিড়লা মিত্তালদের মাসিক আয়ের সঙ্গে খেতে না পাওয়া চাষীর মাসিক আয়ের গড় করে সাব্যস্ত করার চেষ্টা করছেন যে "মোটের উপর" ভারতে খুব একটা অর্থনৈতিক কষ্ট নেই!

13.1.3 Central tendency মাপার তৃতীয় কায়দা

একটা কলেজে কোনো পরীক্ষায় নম্বর উঠেছে $50, 43, 60, 45, 56, 58, 52$, এইরকম। Fig 6 দ্যাখো। এর ভিত্তিতে এমনটা বলা নিশ্চয়ই ঠিক হবে না যে, মোটামুটি সবাই 60-এর ধারে কাছে পেয়েছে। কারণ এখানে 60 আছে একেবারে এক প্রান্তে। তার চেয়ে বরং সবাই মোটামুটি 50-এর ধারে কাছে পেয়েছে বললে কিছুটা ঠিক হত, কারণ 50 মাঝামাঝি জায়গায় আছে। এই "মাঝামাঝি থাকা" দিয়ে আমরা central tendency মাপতে পারি। ধরো এক গুচ্ছ সংখ্যা আছে,

$$x_1, ..., x_n.$$

ওদেরকে ছোটো থেকে বড় সাজাও। তারপর দ্যাখো ঠিক মাঝখানে কোন সংখ্যাটা আসে। তাকে বলে **median.** যেমন $4.5, 3.4, 8.5, 3.5, 9.1$ সংখ্যাগুলোকে ছোটো থেকে বড় সাজালে হয় $3.4, 3.5, 4.5, 8.5, 9.1$. সবচেয়ে মাঝের সংখ্যাটা হল 4.5, সেটাই হল median. এখানে অবশ্য বিজোড়সংখ্যক সংখ্যা ছিল, তাই ঠিক মাঝখানে একটা সংখ্যা পেলাম। যদি জোড়সংখ্যক থাকত, তবে ঠিক মাঝখানে দুটো সংখ্যা আসত, যেমন যদি $4.2, 2.3, 4.2, 5.0$ হত, তবে সাজানোর পর পেতাম $2.3, 4.1, 4.2, 5.0$. সবচেয়ে মাঝে আছে 4.1 আর 4.2. এদের মধ্যে যে কোনো সংখ্যাকেই এখানে median বলা হবে। অনেক সময়ে লোকে এদের average করে $(4.1 + 4.2)/2 = 4.15$-কে median নেয়।
এক কথায় বলতে গেলে, $x_1, ..., x_n$-এর median হল সবচেয়ে মাঝের সংখ্যা, যার বাঁয়ে এবং ডাইনে সমান-সংখ্যক x_i থাকবে।

13.1.4 Central tendency মাপার চতুর্থ কায়দা

আমরা শিখলাম একগুচ্ছ সংখ্যার median কী করে বার করে। এবার আমরা একটা random variable-এর median কাকে বলে শিখব। এখানেও মূল ধারণাটা একই--এটা এমন একটা সংখ্যা m, যাতে $P(X < m) = P(X > m)$ হয়, মানে X-টা m-এর বাঁদিকে পড়ারও যতটা সম্ভাবনা, ডানদিকে পড়ারও ঠিক ততটাই । সেক্ষেত্রে m-কে বলব X-এর median.

Example 6: যদি X তিনটে value নেয়, $-2, 0$ আর 1, আর এদের সমান probability হয় (মানে $\frac{1}{3}$ করে), তবে median হবে 0. এখানে $P(X < 0) = P(X > 0) = \frac{1}{3}$. ■

সমস্যা হল এমন m সব সময়ে নাও পাওয়া যেতে পারে। নীচের অংকে একটা উদাহরণ দিলাম।

Example 7: ধরো একটা random variable-এর distribution হল

x	1	2
$P(X = x)$	0.1	0.9

এমন m দিতে পারো যাতে $P(X < m) = P(X > m)$ হয়?
SOLUTION: অসম্ভব। যদি $m < 1$ নাও, তবে $P(X < m) = 0$ আর $P(X > m) = 1$ হবে। যদি $m = 1$ নাও, তবে $P(X < m) = 0$ আর $P(X > m) = 0.9$ হবে। যদি $1 < m < 2$ নাও, তবে $P(X < m) = 0.1$ আর $P(X > m) = 0.9$ হবে। যদি $m = 2$ নাও তবে $P(X < m) = 0.1$ এবং $P(X > m) = 0$ হবে। যদি $m > 2$ নাও,

তবে $P(X < m) = 1$ আর $P(X > m) = 0$ হবে। ■

তবে random variable-টা যদি continuous হয়, তবে এসব ঝামেলা থাকে না। তখন এরকম m থাকতে বাধ্য। Continuous না হলেও median-এর একটা সংজ্ঞা খাড়া করা যায়, কিন্তু সেটা খুব একটা কাজে লাগে না, তাই সেই প্রসঙ্গে যাচ্ছি না।

13.1.5 দূরত্ব দিয়ে central tendency মাপা

দূরত্বের ধারণা ব্যবহার করে central tendency-কে মাপা যায়। মনে করো একটা শহরে একটা বাস টার্মিনাস বানাতে হবে, যেখান থেকে বাস যাবে শহরের সর্বত্র। এরকম একটা টার্মিনাস স্বাভাবিকভাবেই তুমি শহরের কেন্দ্রে রাখবে। কারণ যদি কোনো এক প্রান্তে রাখো, তবে বিপরীত প্রান্তে পৌঁছতে বেশী তেল পুড়বে। সেই কারণে central tendency বার করার একটা কায়দা হল সেই সংখ্যাটা নেওয়া যেটা কারোর থেকেই খুব বেশী দূরে নেই। একটা random variable-এর বেলায় ব্যাপারটাকে এইভাবে ভাবা যায়। Random variable-টার বিভিন্ন value হল যেন বাসটার বিভিন্ন স্টপ। মনে করো বাস টার্মিনাসটা তৈরী করা হবে a বিন্দুতে। তাহলে X-এ পৌঁছতে বাসটাকে যেতে হবে $|X-a|$ দূরত্ব। অবশ্যই এটা নিজেই একটা random variable. তাই আমরা $E|X-a|$ নিয়ে কাজ করব। এমনভাবে a নেব যাতে $E|X-a|$-টা সবচেয়ে কম হয়। মজার কথা হল, a-টাকে X-এর median নিলেই সেটা হবে। এর প্রমাণটা একেবারে সহজ নয়। যদি X-এর density থাকে, তবে প্রমাণটা রয়েছে নীচের অংকের উত্তরে।

Example 8: If X is a continuous random variable, prove that the first absolute moment of X is minimum when taken about the median.[5] **(2012.2d)**

SOLUTION: এখানে first absolute moment মানে $E(|X-c|)$, যেখানে c কোনো একটা সংখ্যা। দেখাতে হবে যে c যদি X-এর median হয়, তবে এটা minimum হবে।

X বলা আছে continuous. তার মানেই কিন্তু এই নয় যে, X-এর density থাকবে। কিন্তু আমরা সেটা ধরে নেব (কারণ নইলে প্রমাণটা অনেক বেশী জটিল হবে)।

> We assume that X has density $f(x)$.
> Let X have median μ. So
> $$\int_{-\infty}^{\mu} f(x)dx = \int_{\mu}^{\infty} f(x)dx = \frac{1}{2}.$$

কারণ $P(X < \mu) = P(X > \mu)$ আর $P(X = \mu) = 0$, যেহেতু continuous.

> Shall show that
> $$\forall c \neq \mu \quad E|X-\mu| \leq E|X-c|,$$
> Take any $c \neq \mu$.
> <u>Case 1: $c > \mu$:</u>
> $$|X-c| - |X-\mu| = \begin{cases} c-\mu & \text{if } X < \mu \\ c+\mu-2x & \text{if } \mu \leq X < c \\ \mu - c & \text{if } X \geq c \end{cases}$$

কীভাবে তিনটে অংশে ভেঙে লিখেছি দ্যাখো। এইরকমটা করলাম, যাতে কোনো absolute value না থাকে।

So

$$
\begin{aligned}
& E|X-c| - E|X-\mu| \\
= \;& \int_{-\infty}^{\mu} (c-\mu) f(x)dx + \int_{\mu}^{c} (c+\mu-2x) f(x)dx + \int_{c}^{\infty} (\mu-c) f(x)dx
\end{aligned}
$$

যেভাবে তিনটে অংশে ভেঙেছিলাম, সেটা থেকেই এটা পেলাম।

এবার x-ছাড়া জিনিসগুলোকে integration-এর বাইরে নিয়ে আসি--

$$
\begin{aligned}
= \;& (c-\mu) \int_{-\infty}^{\mu} f(x)dx + (c+\mu) \int_{\mu}^{c} f(x)dx - 2 \int_{\mu}^{c} x f(x)dx (\mu - c) \int_{c}^{\infty} f(x)dx \\
= \;& \tfrac{1}{2}(c-\mu) + (c+\mu) \underbrace{P(\mu \le X < c)}_{p,\ \text{say}} - 2 \int_{\mu}^{c} x f(x)dx + (\mu-c) P(X \ge c) \quad \left[\because \textstyle\int_{-\infty}^{\mu} f(x)dx = \tfrac{1}{2}\right] \\
= \;& \tfrac{1}{2}(c-\mu) + (c+\mu)p - 2 \int_{\mu}^{c} x f(x)dx + (\mu-c)\left(\tfrac{1}{2} - p\right) \\
= \;& 2cp - 2 \int_{\mu}^{c} x f(x)dx \\
= \;& 2 \int_{\mu}^{c} (c-x) f(x)dx \ge 0.
\end{aligned}
$$

অন্য কেসটাও একইরকম--

<u>Case 2: $c < \mu$:</u>

$$
|X-c| - |X-\mu| = \begin{cases} c-\mu & \text{if } X < c \\ c+\mu-2x & \text{if } c \le X < \mu \\ \mu - c & \text{if } X \ge \mu \end{cases}
$$

So

$$
\begin{aligned}
& E|X-c| - E|X-\mu| \\
= \;& (c-\mu) P(X < c) + \int_{c}^{\mu} (\mu + c - 2x) f(x)dx + (\mu - c) P(X \ge \mu) \\
= \;& (c-\mu) P(X < c) + (\mu + c) \underbrace{P(c \le X < \mu)}_{p} - 2 \int_{c}^{\mu} x f(x)dx + \tfrac{1}{2}(\mu - c) \\
= \;& (c-\mu)\left(\tfrac{1}{2} - p\right) + (\mu + c)p - 2 \int_{c}^{\mu} x f(x)dx + \tfrac{1}{2}(\mu - c) \\
= \;& 2\mu p - 2 \int_{c}^{\mu} x f(x)dx \\
= \;& 2 \int_{c}^{\mu} (\mu - x) f(x)dx \ge 0.
\end{aligned}
$$

> Thus combining both cases we see that
>
> $$\forall c \neq \mu \quad E|X-\mu| \leq E|X-c|,$$
>
> as required.

■

13.2 Dispersion

Fig 7 আর Fig 8-এ দুই প্রস্থ হাতের লেখা 'a' রয়েছে। এদের মধ্যে একটা লিখেছে একটা বাচ্চা, আর অন্যটা লিখেছে একজন প্রাপ্তবয়স্ক মানুষ। বলতে পারো কোনটা কার লেখা? উত্তরটা বলাই বাহুল্য--প্রথমটা বাচ্চার লেখা, পরেরটা বড়দের। কী করে বুঝলে? প্রথমটায় একই অক্ষর একেক সময়ে একেকরকম হয়ে গেছে, কখনও বড় কখনো ছোটো, কখনও সোজা, কখনও হেলানো। বড়দের হাতের লেখাতেও একটু উনিশ-বিশ হয়, কিন্তু সেটা ছোটোদের মত অতটা নয়। এই যে একই জিনিসের আকার বিভিন্ন সময়ে বিভিন্ন হচ্ছে এই ব্যাপারটাকেই বলে **dispersion.** মনে করতে পারো যেন প্রত্যেকটা অক্ষর লেখা একটা random experiment-এর একেকটা trial. বাচ্চাদের বেলায় outcome-গুলোর মধ্যে তারতম্য বেশী অর্থাৎ dispersion বেশী। বড়দের বেলায় অপেক্ষাকৃত কম। Central tendency-র মত dispersion-ও একটা ধারণা মাত্র, যাকে অংক কষে মাপার নানা কায়দা আছে। এরকম একটা কায়দা শিখব এবার।

বাচ্চাদের হাতের লেখার উদাহরণ নিয়েই কায়দাটা বোঝা যাক। প্রথমে আমরা একটা random variable নেব--অক্ষরের উচ্চতা, মানে

$$X(\omega) = \omega\text{-এর উচ্চতা.}$$

এখানে ω হল outcome, মানে একেকটা হাতে লেখা অক্ষর। এখানে X যে random তার কারণ এই নয় যে, বাচ্চাটা ইচ্ছে করে অক্ষরগুলো লম্বা বেঁটে করেছে। ও চেয়েছিল সবগুলো অক্ষরকে সমান উঁচু রাখতে, কিন্তু হাতের উপর যথেষ্ট নিয়ন্ত্রণ না থাকায় উঁচুনীচু হয়ে গেছে। বাচ্চাটা অক্ষরগুলোকে ঠিক কতটা উঁচু করতে চেয়েছিল সেটা নিশ্চয়ই মাঝামাঝি কিছু একটা হবে, মানে central tendency, যেটাকে মাপার জন্য আমরা এখানে $E(X)$ ব্যবহার করব। এবার দ্যাখো $E(X)$-এর থেকে X কত দূরে যেতে পারে, মানে $X - E(X)$. কোন দিকে যাচ্ছে, তা নিয়ে কোনো মাথাব্যথা নেই (মানে উঁচুর দিকে, নাকি নীচুর দিকে), খালি দূরত্বটাই গুরুত্বপূর্ণ। তাই আমরা $(X - E(X))^2$ দেখব। এইটাও একটা random variable, তাই কখনও বেশী কখনও কম হবে। আমরা জানতে চাই মোটামুটিভাবে তার পরিমাণ কত। আবার সেই "মোটামুটি"-র প্রশ্ন এসে গেল, মানে central tendency. তাই আবার expectation নেব, তাহলে পাব $E(X - E(X))^2$. একে বলে X-এর **variance.** আমরা লিখি $\text{var}(X)$.

13.2.1 Sample variance

$E(X)$-এর সঙ্গে যেমন জড়িত হয়ে আছে তার statistical image মানে sample average বা $\bar{x}$, তেমনি $\text{var}(X)$-এর সঙ্গে জড়িত হয়ে আছে sample variance. বোঝার জন্য আবার আমরা হাতের লেখার উদাহরণে ফিরে যাই।

বাচ্চাদের হাতে লেখা অক্ষরের উচ্চতাকে আমরা X নাম দিয়েছিলাম, যেটা একটা random variable. লক্ষ করো যে X-এর যাবতীয় সম্ভাব্য value বা তাদের probability distribution বিষয়ে আমরা কিছুই জানি না, সুতরাং $\text{var}(X)$-এর সংজ্ঞা লিখতে পারলেও খালি Fig 7-এর ভিত্তিতে আমরা সেটা বার করতে পারব না। যেটা বার করতে পারব সেটা হল $\text{var}(X)$-এর

Fig 7

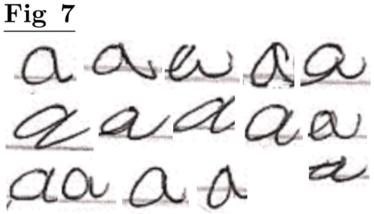
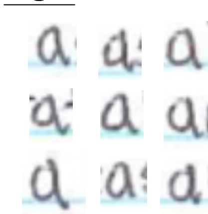


Fig 8



statistical image. তার জন্য প্রথমে Fig 7-এর প্রতিটা অক্ষরের উচ্চতা মেপে ফেলা যাক। এর ফলে আমরা 15-টা সংখ্যা পাবো, যারা X-এর 15-টা realisation:

$$x_1, ..., x_{15}.$$

এদের ব্যবহার করে $E(X)$-এর statistical image হল $\bar{x}$. সুতরাং $(X - E(X))^2$-এর জায়গায় আমরা পাচ্ছি

$$(x_1 - \bar{x})^2, ..., (x_{15} - \bar{x})^2.$$

এর পরের কাজ ছিল $(X - E(X))^2$-এর expectation নেওয়া। এখানে তার জায়গায় গড় বার করব, মানে

$$\frac{1}{n}\sum_1^n (x_i - \bar{x})^2.$$

একে বলে **sample variance,** যাকে অনেক সময়ে S^2 লেখা হয়। এটাই হল $\mathrm{var}(X)$-এর statistical image.

13.2.2 একটু সহজে বার করা

যদি কিছু সংখ্যা দেওয়া থাকে, আর তাদের ভিত্তিতে sample variance বার করতে হয়, তবে তিন রকম পথে এগোনো যায়। প্রথমটা সবচেয়ে কঠিন, তারপরেরটা একটু সহজ, আর শেষেরটা সবচেয়ে সহজ।
প্রথম কায়দাটা হল সরাসরি সংজ্ঞাটা লাগানো। ধরো n-খানা সংখ্যা আছে $x_1, ..., x_n$. তাহলে সরাসরি সংজ্ঞা লাগাতে হলে ধাপগুলো এই--

1. আগে $\bar{x}$ বার করে নাও,
2. সেটাকে প্রতিটা x_i থেকে বিয়োগ করে $x_i - \bar{x}$ পাবে,
3. সেগুলোকে square করো
4. এবং এই square-গুলোর গড় বার করো।

বেশ পরিশ্রমসাধ্য কাজ। ঠিক কতটা পরিশ্রম হিসেব করে দেখি। প্রথম ধাপে $(n-1)$-খানা যোগ এবং একটা ভাগ করতে হবে। দ্বিতীয় ধাপে n-খানা বিয়োগ, তৃতীয় ধাপে n-খানা গুণ, চতুর্থ ধাপটা প্রথম ধাপের মতই, তাই ফের $(n-1)$-টা যোগ এবং একটা ভাগ। তার মানে সব মিলিয়ে যোগ-বিয়োগ-গুণ-ভাগ লাগছে $[(n-1)+1]+n+n+[(n-1)+1] = 4n$-টা। এখানে আমি যোগ-বিয়োগ-গুণ-ভাগ সব একসঙ্গে ধরলাম, কারণ ক্যালকুলেটরে সবগুলোই সমান কঠিন।
এর থেকে একটু সহজ পদ্ধতি পাওয়ার জন্য এইটা লক্ষ কর--

THEOREM

If $x_1, ..., x_n$ are any n numbers then

$$\frac{1}{n}\sum_1^n (x_i - \bar{x})^2 = \frac{1}{n}\sum_1^n x_i^2 - (\bar{x})^2.$$

প্রমাণটা খুবই সহজ। সোজা $(a-b)^2$-এর ফর্মুলা লাগিয়ে দাও--

$$\frac{1}{n}\sum_1^n (x_i - \bar{x})^2 = \frac{1}{n}\sum_1^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2).$$

বাকিটুকু তোমার করা জন্য রেখে দিলাম। Sum-টাকে তিন টুকরো করে ভেঙে নাও, আর i-ছাড়া জিনিসপত্রকে summation-এর বাইরে নিয়ে এসো, তাহলেই পেয়ে যাবে।
এই নতুন ফর্মুলা দিয়ে কী করে সহজে sample variance বার করা যায় দেখি--



Fig 9

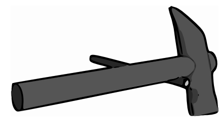


Fig 10

1. প্রথমে আগের মতই $\bar{x}$ বার করতে হবে (এতে $(n-1)$-টা যোগ আর একটা ভাগ লাগবে)।
2. এবার x_i-গুলোকে square করতে হবে (মানে n-খানা গুণ লাগল)।
3. এই square-গুলোর গড় বার করতে হবে (আবার $(n-1)$-টা যোগ আর একটা ভাগ)।
4. $\bar{x}$-কে square করে এই গড়টা থেকে বিয়োগ করতে হবে (তার জন্য খালি একটা গুণ আর একটা বিয়োগ)।

সুতরাং সব মিলিয়ে যোগ-বিয়োগ-গুণ-ভাগ লাগছে $[(n-1)+1]+n+[(n-1)+1]+[1+1]=3n+2$-টা।
সুতরাং এখানে যোগ-বিয়োগ-গুণ-ভাগের সংখ্যা $n-2$-টা কম। যদি n একটা বেশ পেল্লায় গোছের সংখ্যা হয়, বুঝতেই পারছ দ্বিতীয় পদ্ধতিতে খাটনি অনেকটাই বাঁচবে!
এবার বলি সবচেয়ে সহজ কায়দাটার কথা। এটা অবশ্য অংকের কায়দা নয়, স্রেফ ক্যালকুলেটরের কায়দা। অধিকাংশ ছাত্রই আজকাল Casio কোম্পানীর fx82MS বা fx100MS বা সেই জাতীয় ক্যালকুলেটর ব্যবহার করে। সেখানে SD বলে একটা mode হয়। একবার সেই mode-এ ঢুকে তোমাকে খালি পর পর সংখ্যাগুলো টাইপ করতে হবে, প্রত্যেকটা সংখ্যার পরে DT সুইচটা টিপে। তারপর ক্যালকুলেটরটা তোমাকে $\bar{x}$, sample variance ইত্যাদি সব গড়গড় করে বলে দেবে। এমন কি $\sum x_i, \sum x_i^2$ এসবও বলে দেবে। তার জন্য কার পরে কী টিপতে হবে তার খুঁটিনাটি `http://www.isical.ac.in/~arnabc/probstat/`থেকে দেখে নিও।

13.3 Skewness

আমরা expectation আর variance-এর কথা শিখলাম। এবার শিখব skewness বলে একটা জিনিসের কথা। তুমি জানো যে, একটা স্কেলকে তার কেন্দ্রবিন্দুতে আঙুল দিয়ে ব্যালান্স করে রাখা যায় (Fig 9)। সেটা দেখতে বিরাট আহামরি কিছু নয়। কিন্তু একইভাবে একটা হাতুড়িকেও তুমি ব্যালান্স করে রাখতে পারো Fig 10-এর মত, সেটা কিন্তু দেখতে বেশ অদ্ভুত লাগে, কারণ এখানে তোমার আঙুলটা মোটেই কেন্দ্রে নেই, রয়েছে একটা প্রান্তের খুব কাছে। তাও যে হাতুড়িটা পড়ে যাচ্ছে না, কারণ ওই কাছের প্রান্তটা আসলে অনেক বেশী ভারী, ফলে centre of mass (ভরকেন্দ্র) চলে গেছে এক পাশে। সুতরাং এই যে "কেন্দ্র-টা ঠিক কেন্দ্রে নেই"-গোছের ব্যাপারটা হচ্ছে এর মূলে রয়েছে হাতুড়িটার গঠন--এর বিভিন্ন জায়গার ঘনত্ব বিভিন্ন, এবং সেটা কেন্দ্রের দুধারে symmetric-ভাবে নেই। এই symmetry-র অভাবটাকে আমরা বলি skewness. ব্যাপারটা বোঝার জন্য Fig 11, Fig 12 আর Fig 13 দ্যাখো। প্রত্যেকটা ছবিতে একটা করে PDF দেখানো আছে। Fig 11-এর PDF-টা দিব্যি symmetric, কিন্তু বাকি দুজনের ক্ষেত্রে তা নয়। Fig 12-এর বেলায় ডানদিকে লম্বা লেজ বেরিয়েছে, আর Fig 13-এর বেলায় বাঁদিকে। এই asymmetry-র পরিমাণ এবং দিক (অর্থাৎ কত লম্বা লেজ এবং সেটা কোনদিকে), সেটা মাপা যায় 3rd central moment, মানে $E(X-E(X))^3$ দিয়ে। Fig 12-এর বেলায় এটা >0 হবে। তাই আমরা এই PDF-টাকে বলব positively skewed. একই যুক্তিতে Fig 13-এর বেলায় পাবো <0, তাই negatively skewed. যদি Fig 14 দ্যাখো, তবে সেটাও negatively skewed, কিন্তু তাও symmetry-র অনেকটাই কাছাকাছি, তাই এখানেও $E(X-E(X))^3$-টা <0 হবে, কিন্তু শূন্যর থেকে খুব বেশী থাকবে না।
$E(X-E(X))^3$ দিয়ে symmetry-র অভাব মাপা যাচ্ছে কেন। এখানে আসলে 3 সংখ্যাটার দুটো বৈশিষ্ট্য ব্যবহার হচ্ছে--

Fig 11

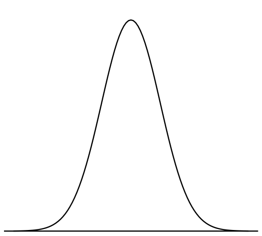


Fig 12

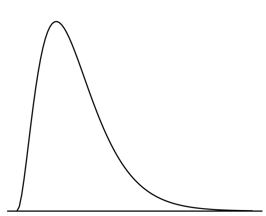


Fig 13

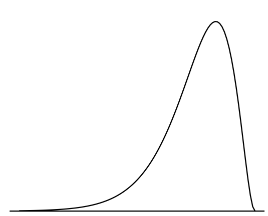
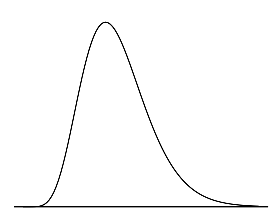


Fig 14



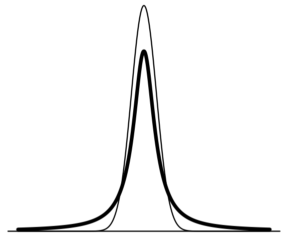


Fig 15

- এক, এটা একটা odd সংখ্যা, তাই $(X-E(X))$-এর যা চিহ্ন, $(X-E(X))^3$-এর চিহ্নও সেটাই থাকছে।

- দুই, $|x|$ যদি $|y|$-এর চেয়ে বড় হয় তবে $|x^3|$ হবে $|y^3|$-এর চেয়ে আরও বেশী বড়, মানে $\frac{|x|}{|y|} > 1$ হলে $\frac{|x^3|}{|y^3|}$ হবে 1-এর চেয়ে আরও বড়।

এর ফলে কী হচ্ছে দ্যাখো। এই কয়টা সংখ্যা নাও, $-3, 1$ আর 2. এদের যোগ করলে 0 হয়। কিন্তু এদের cube-দের যোগ করলে হয় $(-3)^3 + 1^2 + 2^3 = -27 + 1 + 8 = -18 < 0$. এখানে গোড়াতে -3-কে 1 আর 2 মিলে কেটে দিতে পারছিল। কিন্তু cube নেওয়ায় $(-3)^3$ অনেক বেশী negative হয়ে পড়ল। 1^3 আর 2^3 মিলেও তাকে আর সামলাতে পারছে না।

সেই কারণেই যদি একটা distribution-এর কোনো একদিকে একটা সরু লম্বা লেজ থাকে, তবে cube করার পর সেটা আরো বেজায় লম্বা হয়ে উঠে নিজের অস্তিত্ব ঘোষণা করবে, এবং তার ফলে $E(X-E(X))^3$-টাও ঝুঁকে পড়বে সেই দিকে।

13.4 Kurtosis

Central tendency এবং dispersion খুবই গুরুত্বপূর্ণ ধারণা। সেই তুলনায় skewness-এর গুরুত্ব অনেকটাই কম। এবার যেটা নিয়ে আলোচনা করব, kurtosis (কার্টোসিস), সেটার গুরুত্ব তার চেয়েও কম। ব্যাপারটা বোঝার জন্য Fig 15 দ্যাখো। এখানে দুটো PDFদেখানো আছে, দুটোই দিব্যি symmetric, মাঝখানে একটা চূড়া, দুধারে নেমে এসেছে। পার্থক্য হল চূড়াটা কতটা খাড়া সেটা নিয়ে। Fig 15-এ যে PDF-টা সরু দাগে আঁকা সেটা বেশী খাড়া, মোটাটা কম। এই খাড়াই-টাকেই বলে **kurtosis.** এটা মাপা যায় 4th central moment দিয়ে, মানে $E(X-E(X))^4$ দিয়ে। তবে মনে রেখো যে, যদি distribution-টায় একাধিক চূড়া থাকে, বা এলোমেলো গোছের কিছু দেখতে হয়, তবে kurtosis-এর ধারণাটা খুব একটা কাজে আসে না।

Skewness বা kurtosis সাধারণতঃ relative-ভাবে মাপা হয়। এইবার সেইটা নিয়ে আলোচনা করব।

13.5 Relative measures

এবার যেটা আলোচনা করব তার জন্য প্রথমে এই প্রশ্নটার উত্তর দাও দেখি। একজন মানুষের দৈনিক উপার্জন যদি 50 টাকা বাড়ে, তবে সে কতটা খুশি হবে? তুমি হয়তো থতমত খেয়ে ভাববে এ আবার কীরকম প্রশ্ন হল রে বাবা? মানুষটার উপার্জন আগে কত ছিল, সেটা না জেনে উত্তর দেব কী করে? ঠিক কথা, যদি বিল গেট্‌সের উপার্জন দৈনিক 50 টাকা বাড়ে, তবে টেরই পাবে না। অথচ পথের ভিখিরির জন্য সেই পরিমাণটাই অনেক! এইখানেই absolute-এর সঙ্গে relative-এর পার্থক্য। উপার্জনের বৃদ্ধির absolute পরিমাণটা দুক্ষেত্রেই একই, কিন্তু যদি সেটাকে তাদের আগের উপার্জনের অনুপাত হিসেবে দ্যাখো (মানে 50-কে উপার্জনের পরিমাণ দিয়ে ভাগ করে দাও), তবে বিল গেট্‌সের বেলায় সেই relative পরিমাণটা ধূলোয় মিশে যাবে, কিন্তু ভিখিরির ক্ষেত্রে বেশ বড় হবে।

আমরা যে dispersion, skewness আর kurtosis-এর কথা আলোচনা করলাম, তাদের মাপার ক্ষেত্রেও ব্যাপারটা একইরকম। Dispersion-এর কথাই ধরো। সহজ কথায় এর মানে হল "কতটা এদিক ওদিক হতে পারে"। যদি তুমি পৃথিবী থেকে সূর্যের দূরত্ব মাপতে 100 কিলোমিটার এদিক ওদিক করে ফ্যালো, তবে কেউ তা নিয়ে ভুরু কোঁচকাবে না। কিন্তু যদি তোমার প্যান্ট বানাতে গিয়ে দরজি দুটো পায়ের লম্বায় ছয় ইঞ্চি এদিক ওদিক করে দেয়, তাহলেই হয়েছে। সুতরাং বুঝতেই পারছ "কতটা এদিক ওদিক" মাপার সময়ে "এমনিতে মোটামুটি কত" সেটার ধারণাটাও রাখা দরকার, অর্থাৎ dispersion-এর পরিমাণটা central tendency-এর পরিমাণের সাপেক্ষে relative হিসেবে বলাই ভালো। এরকম একটা relative measure of

dispersion হল **coefficient of variation** বা সংক্ষেপে CV, যার সংজ্ঞা--

$$CV = \frac{\text{standard deviation}}{\text{mean}}.$$

অবশ্য এখানে ধরে নিচ্ছি যে mean $\neq 0$. বস্তুতঃ আমরা CV ব্যবহার তখনই করি যখন mean > 0, কারণ negative অনুপাত ব্যাপারটা চিন্তা করার পক্ষে খুব উপাদেয় নয়। লক্ষ কর যে, CV-র কোনো একক নেই। তুমি সেন্টিমিটারেই মাপো আর ইঞ্চিতেই মাপো, CV একই থাকবে।

একইভাবে skewness-এরও relative পরিমাপ সম্ভব। সেখানে ভাগ করা হয় dispersion-এর পরিমাণ দিয়ে। কেন? কারণ skewness মাপছে symmetry-র অভাব। ধরো একটা মূর্তি গড়া হচ্ছে, তার চোখদুটো নাকের দুপাশে symmetric ভাবে থাকতে হবে। কিন্তু মূর্তি গড়তে গিয়ে একটু এদিক ওদিক হয়েই যায়। ধরো এক সেন্টিমিটার কম বেশী হয়ে গেল। সেটা কতটা দৃষ্টিকটু লাগবে? সেটা নির্ভর করছে চোখ দুটোর মোট বিস্তারের উপর। যদি মূর্তিটা একটা বি..শা..ল সাইজের কিছু হয়, যার চোখ দুটোর মোট বিস্তারই হয়তো 3 মিটার, তবে symmetry থেকে এক সেন্টিমিটার বিচ্যুতি কিছুই নয়। কিন্তু যদি তোমার পাড়ার পূজোর প্রতিমায় একই পরিমাণ ভুল হয়, তবে সেটা খুবই দৃষ্টিকটু লাগবে।

এখানে বিস্তার মানে হল dispersion. অর্থাৎ skewness-এর প্রকৃত গুরুত্ব তখনই অনুধাবন করা যাবে যখন dispersion-এর সাপেক্ষে relative-ভাবে দেখা হবে। তাই আমরা skewness-এর relative measure নিতে পারি এইভাবে--

$$\frac{E(X-E(X))^3}{(E(X-E(X))^2)^{3/2}},$$

অর্থাৎ উপরে রয়েছে আমাদের পূর্বপরিচিত skewness-এর পরিমাণ, আর তাকে ভাগ করছি standard deviation-এর cube দিয়ে। এখানে cube নেবার উদ্দেশ্য হল যাতে এটা CV-র মতই একক-নিরপেক্ষ হয়। একে বলে **coefficient of skewness.**

Kurtosis-এর বেলায় আমরা ভাগ করি variance-এর square দিয়ে। উদ্দেশ্য সেই একই, এককটাকে তাড়ানো। এর ফলে পাই

$$\frac{E(X-E(X))^4}{\left(E(X-E(X))^2\right)^2},$$

যার পোশাকি নাম হল **coefficient of kurtosis.** যদি একটা normal distribution নাও (তার mean আর variance যাই হোক না কেন), তবে তার coefficient of kurtosis হবে 3. এটাকে অনেক সময়ে একটা reference value হিসেবে ধরা হয়। যেসব distribution-এর coefficient of kurtosis $= 3$ তাদেরকে বলে **mesokurtic** (অর্থাৎ মাঝারি খাড়াই)। যদি coefficient of kurtosis > 3 হয় তবে বলে **leptokurtic** (মানে অনেকটা খাড়াই), আর যদি coefficient of kurtosis হয় < 3, তবে বলে **platykurtic** (মানে অল্প খাড়াই)। অনেক সময়ে coefficient of kurtosis থেকে 3 বিয়োগ করে যা পড়ে থাকে, তাকে বলে coefficient of excess kurtosis.

এবার একটু হাতেকলমে অংক করি।

Example 9: Find the mean, variance and the coefficient of skewness of the continuous distribution with probability density function given by

$$f(x) = \begin{cases} 1-|1-x| & \text{if } 0<x<2 \\ 0 & \text{otherwise} \end{cases}$$

(2011.2d)

SOLUTION:

$$\begin{aligned} E(X) &= \int_0^2 x(1-|1-x|)dx \\ &= \int_0^1 x(1-|1-x|)dx + \int_1^2 x(1-|1-x|)dx \\ &= \int_0^1 x(1-(1-x))dx + \int_1^2 x(1-(x-1))dx \\ &= \int_0^1 x^2 dx + \int_1^2 (2x-x^2)dx \\ &= \tfrac{1}{3} + \left(3-\tfrac{7}{3}\right) \\ &= 1. \end{aligned}$$

Also

$$\begin{aligned} E(X^2) &= \int_0^2 x^2(1-|1-x|)dx \\ &= \int_0^1 x^2(1-|1-x|)dx + \int_1^2 x^2(1-|1-x|)dx \\ &= \int_0^1 x^2(1-(1-x))dx + \int_1^2 x^2(1-(x-1))dx \\ &= \int_0^1 x^3 dx + \int_1^2 (2x^2-x^3)dx \\ &= \tfrac{1}{4} + \frac{11}{12} \\ &= \tfrac{7}{6}. \end{aligned}$$

So

$$var(X) = E(X^2) - (E(X))^2 = \tfrac{7}{6} - 1^2 = \tfrac{1}{6}.$$

Again,

$$\begin{aligned} E((X-E(X))^3) &= E((X-1)^3) \\ &= \int_0^2 (x-1)^3(1-|1-x|)dx \\ &= -\int_{-1}^1 u^3(1-|u|)du \quad \left[\text{putting } u=1-x\right] \\ &= 0, \end{aligned}$$

since it is the integral of an odd function over an interval symmetric around 0.

■

13.6 Sample moment

আমরা দেখেছি যে $E(X)$-এর statistical image হল sample mean অর্থাৎ $\bar{x}$. একইভাবে $var(X)$-এর statistical image হল sample variance. আন্দাজ করতে পারছ যে, skewness মাপার জন্য যে $E(X - E(X))^3$ ব্যবহার করছিলাম, তার statistical image হবে

$$\frac{1}{n}\sum_{i}^{n}(x_i - \bar{x})^3.$$

মোদ্দা কথা হল যে কোনো moment-এরই statistical image বার করা যায় একইভাবে--খালি expectation-এর জায়গায় গড় বসিয়ে দাও। সুতরাং--

- k-th sample raw moment হল $\frac{1}{n}\sum_1^n x_i^k$,
- আর, k-th sample central moment হল $\frac{1}{n}\sum_1^n (x_i - \bar{x})^k$.

একইভাবে coefficient of variation, coefficient of skewness আর coefficient of kurtosis-এর statistical image-ও পেয়ে যাবে।

DAY 14
Generating functions

14.1 Characteristic function

এইবার আমরা একটা অদ্ভুত জিনিসের কথা শিখব, তার নাম characteristic function.
নাম শুনেই বুঝতে পারছ যে, এটা একটা function. যে কোনো distribution-এর জন্যই একটা characteristic function থাকে। সংজ্ঞাটা চেয়েছে নীচের অংকে।

Example 10: Define characteristic function of a random variable X.[1] **(2004.5b)**

SOLUTION:

> DEFINITION: **Characteristic function**
>
> For any random variable X its **characteristic function** is defined as
> $$Q_X(t)(t) = E(e^{itX}), \quad t \in \mathbb{R}.$$

এখানে ওই i-টা হল গিয়ে complex number-এর i, মানে $i^2 = -1$. শ্রীমান i-এর এই অপ্রত্যাশিত আবির্ভাবে ঘাবড়ে যেও না, ওটা একটা constant বই আর কিছুই নয়। স্বাভাবিকভাবেই এখানে প্রশ্ন ওঠে যে, এই function-টা দিয়ে কী উপকার হয়? সে প্রসঙ্গে এক্ষুণি আসছি, আগে কয়েকটা সহজ উদাহরণ করে সংজ্ঞাটার সঙ্গে বন্ধুত্ব করে নিই-- ■

Example 11: একটা random variable দিচ্ছি X, এইভাবে--প্রথমে একটা কয়েন টস করলাম, head পড়লে $X = 1$ আর tail পড়লে $X = 0$. ধরো বলে দিলাম যে $P(\text{head}) = p$. তবে X-এর characteristic function কী হবে?

Solution: এখানে X খালি দুটো value নেয়, 0 আর 1, যাদের probability যথাক্রমে $1-p$ আর p. সুতরাং যেকোনো function $h(X)$-এর জন্যই $E(h(X))$ হবে

$$E(h(X)) = h(0)(1-p) + h(1)p.$$

সুতরাং আমাদের বেলায়--

$$E(e^{itX}) = (1-p) + e^{it}p.$$

■

আরেকটা উদাহরণ দেখি, এবার density দিয়ে।

Example 12: ধরো একটা random variable আছে X, যার density হল

$$f(x) = \begin{cases} e^{-x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

এর characteristic function বার করো।

Solution:

$$\begin{aligned} E(e^{itX}) &= \int_{-\infty}^{\infty} e^{itx} f(x)dx \\ &= \int_{0}^{\infty} e^{itx} e^{-x} dx \\ &= \int_{0}^{\infty} e^{(it-1)x} dx \end{aligned}$$

এখানে e-র মাথায় complex জিনিসপত্র দেখে ভয় পেও না
$\int e^{\lambda x}dx$ যেভাবে বার করো, সেভাবে করে গেলেই হবে।

$$= \frac{1}{it-1} e^{(it-1)x}\Big|_0^{\infty} = \frac{1}{1-it}.$$

এখানে একটা জিনিস ব্যবহার করলাম--

$$\lim_{x\to\infty} e^{(it-1)x} = 0.$$

এটা হবে কারণ $e^{(it-1)x} = e^{itx} \times e^{-x}$. এদের মধ্যে $|e^{itx}| = 1$, তা সে t আর x যাই হোক না কেন। আর ওদিকে $e^{-x} \to 0$ যেহেতু $x \to \infty$. ■

Exercise 4: Find the characteristic function $Q_x(t)$ of a random variable X having distribution with density

$$f(x) = \begin{cases} 1-|x| & \text{if } |x| < 1 \\ 0 & \text{otherwise} \end{cases}$$

(2009.3c)

Hint:

$$
\begin{aligned}
Q_x(t) &= E(e^{itX}) \\
&= \int_{-\infty}^{\infty} e^{itx} f(x)dx \\
&= \int_{-1}^{1} e^{itx}(1-|x|)dx \\
&= \int_{-1}^{0} e^{itx}(1+x)dx + \int_{0}^{1} e^{itx}(1-x)dx \\
&\quad \text{দুভাগে ভেঙে নিলাম, যাতে } |x| \text{ না থাকে।} \\
&= \int_{-1}^{1} e^{itx}dx - \int_{0}^{1} xe^{itx}dx + \int_{-1}^{0} xe^{itx}dx \\
&= \frac{1}{it}e^{itx}\Big|_{-1}^{1} - \int_{0}^{1} xe^{itx}dx + \int_{-1}^{0} xe^{itx}dx
\end{aligned}
$$

■

14.1.1 করে কী লাভ?

মোটামুটিভাবে characteristic function-এর তিনটে গুণ আছে। একে একে আলোচনা করি।
প্রথম গুণটা বলতে গিয়ে এটাও বলা হয়ে যাবে যে, $E(e^{itX})$-কে কেন characteristic function বলা হয়। কোনো দুটো distribution-এর characteristic function কখনও এক হতে পারে না। অনেকটা আঙুলের ছাপের মত ব্যাপার আর কি, প্রত্যেকটা মানুষেরই আঙুলের ছাপ হয়, এবং সেটা অন্য কোনো মানুষের ছাপের সঙ্গে মেলে না।

THEOREM

Every probability distribution (defined on $\mathbb{R}$) has a unique characteristic function. The characteristic functions of two different distributions are always different.

এই theorem-টা কাজে লাগানোর জন্য প্রস্তুতি হিসেবে যেটা করতে হয় সেটা হল বিভিন্ন standard distribution-এর characteristic function-গুলোর একটা তালিকা মাথায় রাখা। তাহলে সেই তালিকা দেখে নিলেই আমরা distrbution দেওয়া থাকলে তার characteristic function, এবং characteristic function দেওয়া থাকলে তার distribution বার করে ফেলতে পারব। এরকম একটা ছোটো তালিকা দেব পঞ্চম অধ্যায়ে।
এবার আসি দ্বিতীয় গুণের কথায়। সেটা বোঝার জন্য মনে কর দুটো random variable আছে, X, Y, যারা independent. এবার $Z = X + Y$ নিলাম। তাহলে X আর Y-এর distribution জানা থাকলে তা থেকে Z-এর distribution বার করে দেওয়া যায় নীচের theorem-টা ব্যবহার করে।

THEOREM

Let X, Y be two independent random variables with characteristic functions $Q_X(t)$ and $Q_Y(t)$, respectively, then the characteristic function of $Z = X + Y$ is

$$Q_Z(t) = Q_X(t) \times Q_Y(t).$$

এর প্রমাণটা আসে এই জিনিসটা থেকে--যদি $h_1(X)$ আর $h_2(Y)$ যে কোনো দুটো function হয় যাতে $E(h_1(X)), E(h_2(Y))$ এবং $E(h_1(X)h_2(Y))$, এরা সবাই exist করে, তবে যেহেতু X, Y হল independent, তাই $E(h_1(X)h_2(Y)) = E(h_1(X)) \times E(h_2(Y))$ হবে। এর প্রমাণটা আমরা ষষ্ঠ অধ্যায়ে গিয়ে শিখব। এখানে $h_1(x) = h_2(x) = e^{itx}$ নিলেই theorem-টা পেয়ে যাবে।

তৃতীয় আরেকটা গুণও আছে, যেটা আগের দুটোর চেয়ে কম গুরুত্বপূর্ণ হলেও উল্লেখের দাবী রাখে। গুণটা হল--একটা random variable-এর characteristic function জানা থাকলেই তার যাবতীয় raw moment বার করা যায়, যদি অবশ্য raw moment-গুলো আদৌ exist করে। এর জন্য প্রথমে e^{itX}-কে infinite series আকারে লেখো--

$$e^{itX} = 1 + itX + \frac{i^2t^2X^2}{2!} + \frac{i^3t^3X^3}{3!} + \frac{i^4t^4X^4}{4!} + \cdots.$$

তারপর দুধারের expectation নাও। এইখানে আমরা দুটো জিনিস ধরে নেব, মানে দুটো assumption করব--

- এক, সব $k \in \mathbb{N}$-এর জন্যই $E(X^k)$-টা exist করে।
- দুই, expectation-টা term-by-term নেওয়া যায়, মানে--

$$E\left(1 + itX + \frac{i^2t^2X^2}{2!} + \frac{i^3t^3X^3}{3!} + \cdots\right) = 1 + itE(X) + \frac{i^2t^2E(X^2)}{2!} + \frac{i^3t^3E(X^3)}{3!} + \cdots.$$

এর মধ্যে প্রথম assumption-টা সবসময়ে ঠিক নাও হতে পারে (যদি না হয়, তবে যেটা বলতে যাচ্ছি সেটা সেইরকম ক্ষেত্রে খাটবে না)। প্রথমটা যদি ঠিক হয়, তবে দ্বিতীয় assumption-টা সবসময়েই খাটবে, সেটা প্রমাণ করা যায়, যদিও প্রমাণটা আমাদের বইয়ের পাল্লার বাইরে। যাই হোক, এই দুটো জিনিস ধরে নিলে পাচ্ছি

$$Q_X(t) = 1 + itE(X) + \frac{i^2t^2E(X^2)}{2!} + \frac{i^3t^3E(X^3)}{3!} + \frac{i^4t^4E(X^4)}{4!} + \cdots.$$

তার মানে $Q_X(t)$ হল t-এর একটা power series যেখানে t^k-র coefficient হচ্ছে $i^kE(X^k)/k!$. সুতরাং power series-টা একবার পেয়ে গেলে যাবতীয় $E(X^k)$ বার করে ফেলা যাবে। কাজটা দুইভাবে করা যায়--

- এক, power series-টার general term-টা লিখে সেখান থেকে t^k-র coefficient-টা বার করে, তাকে $i^k/k!$ দিয়ে ভাগ করে দেওয়া।
- দুই, $Q_X(t)$-কে k-বার differentiate করে $t = 0$ বসিয়ে $Q_X^{(k)}(0)$ বার করো। তাহলে $E(X^k)$ হবে $Q_X^{(k)}(0)/i^k$. এটা কেন কাজ করে দ্যাখো। প্রথমবার differentiate করলে পাবে

  $$Q'_X(t) = iE(X) + \frac{2ti^2E(X^2)}{2!} + \frac{3t^2i^3E(X^3)}{3!} + \frac{4t^3i^4E(X^4)}{4!} + \cdots.$$

  লক্ষ কর একেবারে প্রথমে আছে $iE(X)$, তারপর থেকে বাকি সব term-এই অন্তত একটা t আছে। তাই $t = 0$ বসালে ওরা মারা পড়বে, খালি $iE(X)$-টাই পড়ে থাকবে, তাকে i^1 দিয়ে ভাগ করলেই $E(X)$ পেয়ে যাবে। একইভাবে, দ্বিতীয়বার differentiate করার পর প্রথমে থাকবে $i^2E(X^2)$, পরের সব term-এ t থাকবে। সুতরাং $t = 0$ বসালে পড়ে থাকবে খালি $i^2E(X^2)$. তাকে i^2 দিয়ে ভাগ করলেই $E(X^2)$ পাওয়া যাবে।

এই যুক্তিতে আমরা ধরে নিয়েছিলাম যে সবগুলো raw moment-ই finite. যদি random variable-টা কেবল finite-সংখ্যক value নিতে পারে, তবে সেটা হতে বাধ্য। সেই নিয়েই নীচের অংকটা।

Example 13: Let X be a random variable taking values in a finite set. Then prove that the moments $E(X^n)$ of X are given by

$$E(X^n) = \frac{1}{i^n}\frac{d^n}{dt^n}Q_x(0),$$

where $Q_x(t)$ is the characteristic function of X, and $i = \sqrt{-1}$.[5] **(2010.3d)**

SOLUTION:

Let X take values in a finite set $\{x_1, ..., x_p\}$.
Then its characteristic function is

$$Q_x(t) = \sum_{k=1}^{p} e^{itx_k} P(X = x_k).$$

We know that for any constant α

$$\frac{d^n e^{\alpha t}}{dt^n} = \alpha^n e^{\alpha t}.$$

কারণ প্রতিবার differentiate করলে একটা করে α বেরিয়ে আসবে।

So

$$\begin{aligned}
\frac{d^n}{dt^n} Q_x(t) &= \frac{d^n}{dt^n} \sum_{k=1}^{p} e^{itx_k} P(X = x_k) \\
&= \sum_{k=1}^{p} \frac{d^n}{dt^n} e^{itx_k} P(X = x_k) \\
&= \sum_{k=1}^{p} (ix_k)^n e^{itx_k} P(X = x_k) \\
&= i^n \sum_{k=1}^{p} x_k^n e^{itx_k} P(X = x_k).
\end{aligned}$$

Hence

$$\frac{d^n}{dt^n} Q_x(0) = i^n \sum_{k=1}^{p} x_k^n P(X = x_k) = i^n E(X^n).$$

So

$$E(X^n) = \frac{1}{i^n} \frac{d^n}{dt^n} Q_x(0),$$

as required.

■

14.2 Moment generating function

Characteristic function জিনিসটা বেশ কাজের বটে, কিন্তু হঠাৎ complex number-এর আবির্ভাবে অনেকে অস্বস্তি বোধ করে। তাদের জন্য characteristic function-এর একটা হোমিওপ্যাথিক সংস্করণ আছে, তাকে বলে moment generating function. সংজ্ঞাটা চেয়েছে নীচের অংকে। সেটা দেখলেই বুঝবে যে, ব্যাপারটা একেবারেই characteristic function-এর মতই, খালি i-টুকু বাদে।

Example 14: Define moment generating function (m.g.f.) of a random variable X.[2] **(2005.4a)**

SOLUTION:

> DEFINITION: **Moment generating function (MGF)**
>
> By the moment generating function of a random variable X we mean a function $\phi(t)$ defined as
>
> $$\phi(t) = E(e^{tX}),$$
>
> for all $t \in \mathbb{R}$ for which the expectation exists finitely.

■

কিন্তু ওই i-টুকু বাদ দেওয়ায় moment genrating function জিনিসটা characteristic function-এর চেয়ে অনেক দুর্বল হয়ে পড়েছে।

- প্রথমতঃ এটা সবসময়ে exist করে না, মানে এমন distribution আছে যার জন্য moment genrating function-টা exist-ই করে না, বা করলেও পুরো $\mathbb{R}$-এর উপর করে না। সেই জন্যেই সংজ্ঞার মধ্যে আলাদা করে "for which the expectation exists finitely" লিখে নিতে হয়েছে।
- তার উপর আবার এমনটা সম্ভব যে, দুটো আলাদা distribution-এর moment genrating function একই। সুতরাং কেবলমাত্র moment genrating function দেখেই distribution-টাকে সব সময়ে সনাক্ত করা যায় না।

এই দুই কারণে moment genrating function-এর জনপ্রিয়তা characteristic function-এর তুলনায় অনেক কম। এবার হাতেকলমে একটা উদাহরণ করা যাক।

Example 15: If a random variable X has probability density function

$$f(x) = \frac{1}{2}e^{-|x|}, \quad -\infty < x < \infty,$$

show that the moment generating function $\phi(t) = E[e^{tX}]$ is given by

$$\phi(t) = \frac{1}{1-t^2}.$$

[3] **(2010.2d)**
SOLUTION: অংকটায় একটা ভুল আছে। কোন কোন t নিয়ে কাজ হচ্ছে, সেটা বলে দেয় নি। যেমন ধরো $t = -1$ বা $t = 1$ যে নেওয়া যাবে না, বুঝতেই পারছ, কারণ সেক্ষেত্রে তলার $1 - t^2$-টা 0হয়ে যাবে। আসলে এখানে $t \in (-1, 1)$ নিয়ে কাজ করতে হবে। সেটা বলে দেওয়া উচিত ছিল। এই শর্তটা অংকটা কষতে গেলে আপনেই বেরিয়ে আসবে।

$$\begin{aligned}
\phi(t) &= E[e^{tX}] \\
&= \frac{1}{2}\int_{-\infty}^{\infty} e^{-|x|}e^{tx}dx \\
&= \frac{1}{2}\left[\int_{-\infty}^{0} e^{x(t+1)}dx + \int_{0}^{\infty} e^{(t-1)x}dx\right]
\end{aligned}$$

এখানে $|x|$ ছিল বলে দুভাগে ভেঙে নিলাম, যাতে $|x|$ আর না থাকে।

$$= \frac{1}{2}\left[\int_{0}^{\infty} e^{-(t+1)y}dy + \int_{0}^{\infty} e^{(t-1)x}dx\right] \quad \left[\text{putting } y=-x \text{ in the first integral}\right]$$

সুতরাং দুটো integral-ই এখানে দেখতে $\int_0^\infty e^{\lambda x}dx$-এর মত হয়েছে। আমরা জানি যে, এরকম integral-কে finite হতে হলে $\lambda < 0$ হতে হয় (নইলে $e^{\lambda x} \to \infty$ হয়ে যাবে, যখন $x \to \infty$ হবে)। যদি $\lambda < 0$ হয়, তবে integral-টা হবে $-\frac{1}{\lambda}$.

We know that $\int_0^\infty e^{\lambda x}dx$ exists finitely iff $\lambda < 0$. In this case the value is $-\frac{1}{\lambda}$.
So

$$\phi(t) = \frac{1}{2}\left[-\frac{1}{-(t+1)} - \frac{1}{t-1}\right] = \frac{1}{1-t^2},$$

when $t \in (-1,1)$, as required.

এই $t \in (-1,1)$-টা কী করে এল, বুঝলে তো? প্রথম integral-টার জন্য দরকার $-(t+1) < 0$, মানে $t > -1$. দ্বিতীয় integral-টার জন্য চাই $t - 1 < 0$, অর্থাৎ $t < 1$. দুয়ে মিলে হল $-1 < t < 1$. ■

Characteristic function-এর চেয়ে দুর্বল হলেও, moment generating function-এর বাকি গুণগুলো একই। যদি X, Y দুটো random variable হয় যারা independent, এবং যাদের mgf হল যথাক্রমে $\phi_X(t)$ আর $\phi_Y(t)$ তবে $X+Y$-এর mgf হবে $\phi_X(t)\phi_Y(t)$.
অন্য গুণটা এই যে, mgf জানা থাকলে যাবতীয় raw moment বার করা যায়। বস্তুতঃ সেই জন্যেই moment generating function নামটা হয়েছে। কায়দাটা একেবারেই charactristic function-এর বেলায় যেরকম ছিল সেরকমই, প্রথমে e^{tX}-কে infinite series আকারে লেখো--

$$e^{tX} = 1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \frac{t^4X^4}{4!} + \cdots.$$

তারপর দুধারের expectation নাও। এইখানে আমরা যথারীতি দুটো assumption করেছি--

- এক, সব $k \in \mathbb{N}$-এর জন্যই $E(X^k)$-টা exist করে।
- দুই, expectation-টা term-by-term নেওয়া যায়, মানে

$$E\left(1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \frac{t^4X^4}{4!} + \cdots\right) = 1 + tE(X) + \frac{t^2E(X^2)}{2!} + \frac{t^3E(X^3)}{3!} + \frac{t^4E(X^4)}{4!} + \cdots$$

হয়।

এর মধ্যে প্রথম assumption-টা সবসময়ে ঠিক নাও হতে পারে (যদি না হয়, তবে যেটা বলতে যাচ্ছি সেটা সেইরকম ক্ষেত্রে খাটবে না)। প্রথমটা ঠিক হলে, দ্বিতীয় assumption-টা সবসময়েই খাটবে, ঠিক যেমন characteristic function-এর বেলায় বলেছিলাম। যাই হোক, এই দুটো জিনিস ধরে নিলে পাচ্ছি

$$\phi_X(t) = 1 + tE(X) + \frac{t^2E(X^2)}{2!} + \frac{t^3E(X^3)}{3!} + \frac{t^4E(X^4)}{4!} + \cdots.$$

তার মানে $\phi_X(t)$ হল t-এর একটা power series যেখানে t^k-র coefficient হচ্ছে $E(X^k)/k!$. সুতরাং power series-টা একবার পেয়ে গেলে যাবতীয় $E(X^k)$ বার করে ফেলা যাবে। Characteristic function-এর মতই কাজটা দুইভাবে করা যায়।

- এক, power series-টার general term-টা লিখে সেখান থেকে t^k-র coefficient-টা বার করে, তাকে $k!$ দিয়ে গুণ করে দেওয়া।
- দুই, $\phi_X(t)$-কে k-বার differentiate করে $t = 0$ বসিয়ে দিলে। এটা কেন কাজ করে দ্যাখো। প্রথমবার diffrentiate করলে পাবে

$$\phi'_X(t) = E(X) + \frac{2tE(X^2)}{2!} + \frac{3t^2E(X^3)}{3!} + \frac{4t^3E(X^4)}{4!} + \cdots.$$

  লক্ষ কর একেবারে প্রথমে আছে $E(X)$, তারপর থেকে বাকি সব term-এর অন্ততঃ একটা t আছে। তাই $t = 0$ বসালে ওরা মারা পড়বে, খালি $E(X)$-টাই পড়ে থাকবে। একইভাবে দ্বিতীয়বার differentiate করার পর প্রথমে থাকবে $E(X^2)$, পরের সব term-এ t থাকবে, এইরকম।

দুটো অংক করে সড়গড় হয়ে নেওয়া যাক।

Example 16: Find the moment generating function of uniform distribution in $(-a, a)$ $a > 0$.

Hence find the moments of order k for $k = 1, 2, 3, 4$.[5] **(2008.2e)**

SOLUTION:

> The uniform distribution over $(-a, a)$ has PDF
>
> $$f(x) = \begin{cases} \frac{1}{2a} & \text{if } x \in (-a, a) \\ 0 & \text{otherwise} \end{cases}$$
>
> Clearly $\phi(0) = 1$.

কারণ $\phi(0) = E(e^{0\times X}) = E(1) = 1$.

For $t \neq 0$,

$$\begin{aligned}\phi(t) &= E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)dx \\ &= \frac{1}{2a}\int_{-a}^{a} e^{tx}dx \\ &= \frac{1}{2at}e^{tx}\Big|_{-a}^{a}\end{aligned}$$

লক্ষ কর কীভাবে নীচের তলায় একটা t এসে গেল।
সেই জন্যেই $t=0$ কেসটা আলাদা করে করতে হয়েছিল।

$$= \frac{e^{at}-e^{-at}}{2at}.$$

We know that

$$\frac{1}{2}(e^x - e^{-x}) = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots.$$

So

$$\phi(t) = 1 + \frac{a^2t^2}{3!} + \frac{a^4t^4}{5!} + \cdots \text{ for } t \in \mathbb{R}.$$

লক্ষ কর যে, $t=0$ কেসটাও কিন্তু এখানে দিব্যি ঢুকে গেছে!

Since X is bounded, hence $\forall k \in \mathbb{N}\quad E(X^k)$ exists finitely. So

$$\phi(t) = \sum_{k=0}^{\infty} \frac{E(X^k)}{k!}t^k.$$

Matching coefficients, we see $E(X^k)=0$ if k is odd, and $E(X^k)=\frac{a^k}{k}$ if k is even. So

$$\begin{aligned}E(X) &= 0\\ E(X^2) &= \frac{a^2}{2}\\ E(X^3) &= 0\\ E(X^4) &= \frac{a^4}{4}.\end{aligned}$$

■

Example 17: If a random variable X has probability density function

$$f(x) = \frac{1}{2}e^{-|x|}, \quad -\infty < x < \infty,$$

show that the moment generating function $\phi(t) = E[e^{tX}]$ is given by

$$\phi(t) = \frac{1}{1-t^2}.$$

Hence find $var(X)$.[5] **(2010.2d)**

SOLUTION: প্রথম অংশ আজকেই একটু আগেই করলাম। এবার দ্বিতীয় অংশটা করি--

> Here $\phi(t) = \frac{1}{1-t^2}$ for $t \in (-1, 1)$.
> So $\phi'(t) = \frac{2t}{(1-t^2)^2}$.
> Also $\phi''(t) = \frac{2(1-t^2)^2+8t^2(1-t^2)}{(1-t^2)^4}$.

লক্ষ কর যে, আমরা পুরোটাকে কেটেকুটে গুছিয়ে লিখি নি, কারণ এক্ষুণি আমরা $t = 0$ বসাব, তাতে এমনিতেই অনেক কিছু বাদ যাবে।

> Hence
> $$E(X) = \phi'(0) = \left.\frac{2t}{(1-t^2)^2}\right|_0 = 0.$$
> Also
> $$E(X^2) = \phi''(0) = \left.\frac{2(1-t^2)^2 + 8t^2(1-t^2)}{(1-t^2)^4}\right|_0 = 2.$$
> So $var(X) = E(X^2) - (E(X))^2 = 2 - 0^2 = 2$.

■

Chapter V
Standard distribution families

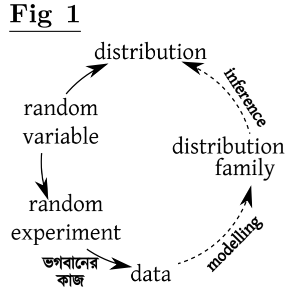
DAY 15
Standard discrete distribution families (part 1)

এবার যা করতে চলেছি, তার জন্য দ্বিতীয় অধ্যায় থেকে কয়েকটা জিনিস একটু মনে করে নিলে সুবিধা হবে। Statistics-এর কারবার হল random variable নিয়ে। এরকম variable কখন ঠিক কী value নেবে তার স্থিরতা থাকে না। সুতরাং কোনো random variable-এর খোঁজ পেলেই statistician-দের আগ্রহ থাকে তার probability distribution-টা বার করতে, যা দিয়ে সেই random variable-এর যাবতীয় event-এর probability জানা যায়। যেহেতু random variable-রা "আমার distribution অমুক" বলে কপালে সাইনবোর্ড এঁটে ঘোরে না, তাই এই কাজটা একটু ঘুর পথে করতে হয়। সেটা এইরকম--

> আমরা জানি যে, যে কোনো random variable-এর পিছনেই থাকে একটা random experiment. সেটা বার বার independently করলে random variable-টার প্রচুর realisation পাওয়া যায়, এবং সেইগুলোই data হিসেবে statistician-এর হাতে এক মাত্র সম্বল হয়। সুতরাং statistician-এর কাজ হয় এই data থেকে distribution-এ পৌঁছোনো। এই কাজটা সাধারণতঃ দুই ধাপে করা হয়। প্রথমে histogram বা barplot দেখে distribution-এর আদলটা ঠিক করে নেওয়া হয়। সেই একই আদলের সব distribution-দের এক সঙ্গে নিয়ে একটা family কল্পনা করা হয়। এই হল প্রথম ধাপ। এর পর দ্বিতীয় ধাপে সেই family-র কোন্ সদস্যের সঙ্গে আমাদের data সবচেয়ে বেশী খাপ খায় সেটা নিয়ে মাথা ঘামানো হয়। পুরো জিনিসটা Fig 1-এ দেখিয়েছি। ছবিটা তুমি দ্বিতীয় অধ্যায়ে আগেই দেখেছো।

বাস্তবে দেখা যায় যে, কিছু কিছু distribution family বার বার করে আসে। এদের প্রত্যেকটাকে বলে একেকটা standard distribution family. এরকম বহু family-র একটা লম্বা তালিকা তৈরী করা হয়েছে, এদের চেহারার আদলগুলো চিনে রাখাটা বড়ই দরকারী। ঠিক যেমন একজন সঙ্গীতশিল্পীর পক্ষে বহুলব্যবহৃত রাগগুলোর বিষয়ে ধারণা থাকা দরকার, যাতে একটা গান শুনেই মোটামুটি বলে দেওয়া যায় কোন রাগে সুরটা বাঁধা হয়েছে। একবার যেই এই ধারণাটা পেয়ে গেলে তখন বাকী খুঁটিনাটিগুলো বার করে ফেলা সহজ হয়।
এইবার আমরা এরকম কয়েকটা distribution family-র সঙ্গে পরিচিত হব।

Fig 1



15.1 Bernoulli distribution

আমরা প্রথম যে family-টা নিয়ে কাজ করব, সেটা হল Bernoulli family. ধরো একটা সিনেমার টিকিট আছে সেটা দুই বন্ধুর মধ্যে একজনকে দিতে হবে, যাতে কেউ মনে কষ্ট না পায়। একটা পথ হল কয়েন টস করে ঠিক করা। যদি হাতের কাছে কোনো কয়েন না থাকে তখন এইভাবে করা যায়, একটা কাগজে দুই বন্ধুর নাম লিখে, ভালো করে মিশিয়ে দিয়ে চোখ বুঁজে একটা টেনে নাও। বুঝতেই পারছ যে, এই random experiment-টা বর্ণনার দিক থেকে কয়েন টসের থেকে সম্পূর্ণ আলাদা হলেও, কাজটা একই করছে। কারণ এখানেও ঠিক দুটো outcome-ই আছে, যাদের probability সমান সমান (মানে $\frac{1}{2}$ করে)। এই রকম যে কোনো random experiment যার ঠিক দুটো outcome-ই আছে (অর্থাৎ sample space-এর সাইজ হল 2), তাকে বলে একটা Bernoulli random experiment. অবশ্য সব সময়েই যে দুটো outcome-র probability সমান হবেই এমন কথা নেই। আমরা সাধারণতঃ outcome দুটোর নাম দিই 0 আর 1 (অনেক সময়ে লোকে 0-র জায়গায় failure, আর 1-এর জায়গায় success বলতে ভালোবাসে)।
সুতরাং আমরা পেয়ে গেলাম একটা family, যার সংজ্ঞা এইরকম--

DEFINITION: **Bernoulli distribution**

A random variable X is said to have a **Bernoulli distribution** if it has PMF

$$f(x) = \begin{cases} 1-p & \text{if } x=0 \\ p & \text{if } x=1 \\ 0 & \text{otherwise} \end{cases}$$

Here $p \in (0,1)$ is a parameter. Such a random variable can take only two values 0 and 1. We write $X \sim \text{Bernoulli}(p)$.

Example 1: Define Bernoulli sequence of trials.[1] **(2005.3a)**

SOLUTION:

DEFINITION: **Bernoulli sequence of trials**

If a random experiment has only two possible outcomes (say, 0 and 1), and it is repeated independently, then the resulting sequence of trials is called a Bernoulli sequence of trials.

■

Exercise 1: Define Bernoulli trials.[1] **(2004.2a)**
HINT:
আগের অংকটাই। ■

Example 2: If p is the probability of success in a single trial, find the probability of $r(\leq n)$ successes in a Bernoulli sequence of n trials.[4] **(2005.3a)**
SOLUTION:

> In a sequence of n Bernoulli trials (with outcomes "failure" and "success") exactly r successes can occur in $\binom{n}{r}$ ways.

একটা উদাহরণ নিলে সুবিধা হবে। ধরো $n = 4$ আর $r = 2$. যদি "failure"-কে F আর "success"-কে S লিখি তবে $n = 4$-টে trial-এর মধ্যে ঠিক $r = 2$-টো success হতে পারে $\binom{n}{r} = {}^4C_2 = 6$-ভাবে--

SSFF, SFSF, SFFS, FSSF, FSFS, FFSS.

> Each such sequence consists of r successes and $n-r$ failures. Since the trials are independent the probability of each such sequence is
>
> $$(P(\text{success}))^r(P(\text{failure}))^{n-r}.$$
>
> So the required probability is
>
> $$\binom{n}{r}(P(\text{success}))^r(P(\text{failure}))^{n-r}, \quad r = 0, 1, ..., n.$$

■

এই যে হিসেবটা করলাম, এ থেকেই চলে আসে আরেকটা গুরুত্বপূর্ণ family-র কথা--binomial distribution, যার আলোচনা আমরা এবার করব।

15.2 Binomial distribution

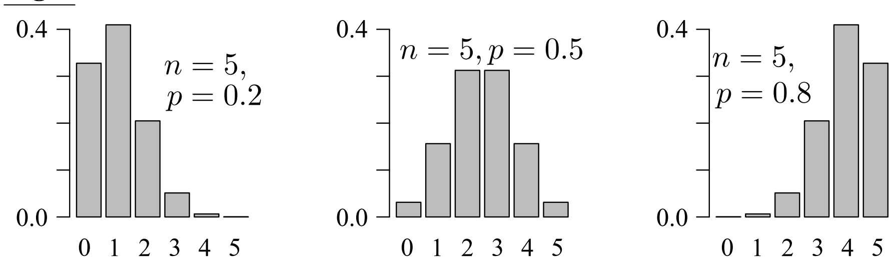
DEFINITION: Binomial distribution

A random variable X is said to have a **Binomial distribution** if it has PMF

$$f(x) = \begin{cases} \binom{n}{x}p^x(1-p)^{n-x} & \text{if } x \in \{0, 1, 2, ..., n\} \\ 0 & \text{otherwise} \end{cases}$$

Here $n \in \mathbb{N}$ and $p \in (0, 1)$ are two parameters. Such a random variable can take the values $0, 1, ..., n$. We write $X \sim \text{Binomial}(n, p)$.

Fig 2



কিছু binomial distribution-এর PMF-দের ছবি রয়েছে Fig 2-এ।
Binomial-কে Bernoulli-দের যোগফল হিসেবে ভাবা যায়। কী করে বলি। মনে করো $X \sim \text{Binomial}(n, p)$. তাহলে এইভাবে ভাবো--একটা কয়েন আছে যার $P(\text{head}) = p$, সেটাকে n-বার independent-ভাবে টস করা হয়েছে। তাহলে X হল মোট head-এর সংখ্যা। এবার প্রত্যেকটা টসে কী পড়ছে সেটার হিসেব রাখা যাক এইভাবে--$Y_i = 1$ নেব যদি i-th টসে head পড়ে, নইলে $Y_i = 0$. তাহলে বুঝতে পারছ নিশ্চয়ই যে Y_i-দের distribution হবে Bernoulli(p), এবং ওরা পরস্পরের সঙ্গে independent. একটু চিন্তা করলেই বুঝবে যে, $X = Y_1 + \cdots + Y_n$. কারণ, প্রত্যেকটা Y_i-ই হল 0 বা 1. যারা 0, তাদের কোনো অবদান থাকবে না যোগফলটায়, সুতরাং যোগফলটা খালি গুণছে কতগুলো $Y_i = 1$ আছে। এ থেকেই আসছে নীচের theorem-টা--

THEOREM

If $Y_1, ..., Y_n$ are IID Bernoulli(p) random variables, then $\sum_1^n Y_i \sim \text{Binomial}(n, p)$.

এই theorem-টাকে এইভাবে মনে রেখো--

> যদি কোনো random experiment থাকে, যার কেবলমাত্র দুটোই outcome (ধরো success আর failure), আর আমরা সেই random experiment-টাকে n-বার independent-ভাবে করি, তবে মোট success-এর সংখ্যার distribution হবে Binomial(n, p), যেখান $p = P(\text{success})$.

একটা প্রয়োগ দেখা যাক।

Example 3: Assuming that A and B are two equally strong chess players, which of the following events is more probable?

1. A beats B in exactly 3 out of 4 games.
2. A beats B in exactly 5 out of 8 games.

[5] **(2010.1b)**
SOLUTION: যেহেতু A আর B দুজনেই সমান ওস্তাদ খেলোয়াড় তাই আমরা ধরে নেব যে, খেলার হার-জিত নির্ভর করছে ভাগ্যের উপর, এবং ভাগ্যদেবীর কোনো পক্ষপাত নেই। আরও ধরে নেব যে, ড্র হতে পারে না। অতএব--

> We assume that in each game draw is not possible, and that A beats B with probability $\frac{1}{2}$. Also we assume that the outcomes of the games are mutually independent.

তার মানে এখানে প্রতিটা খেলা হল random experiment-টার একেকটা trial. এখানে দুটো মাত্র outcome, হয় A জিতবে, নয়তো B জিতবে (যেহেতু ড্রয়ের সম্ভাবনা ধরা হচ্ছে না)। আমরা A-র জয়ের সংখ্যা নিয়ে মাথা ঘামাব, তাই "A-র জয়"-কে success বলব।

> Let X = the total number of wins for A out of a sequence of n games.
> Then $X \sim \text{Binomial}\left(n, \frac{1}{2}\right)$.
> So
> $$P(X = x) = \binom{n}{x} \left(\tfrac{1}{2}\right)^x \left(1 - \tfrac{1}{2}\right)^{n-x} = \binom{n}{x} \left(\tfrac{1}{2}\right)^n$$
> for $x = 0, 1, ..., n$.

Thus the probability of the first event is

$$\binom{4}{3}\left(\tfrac{1}{2}\right)^4 = \tfrac{1}{4},$$

and the probability of the second event is

$$\binom{8}{5}\left(\tfrac{1}{2}\right)^8 = \binom{8}{3}\left(\tfrac{1}{2}\right)^8 = \frac{8\times 7\times 6}{3!}\left(\tfrac{1}{2}\right)^8 = \frac{7}{32} < \frac{1}{4}.$$

So the first event is the more probable of the two.

■

Example 4: A and B play a game in which A's chance of winning is $\frac{2}{3}$. In a series of 8 games what is the chance that A will win at least 6 games?[3] **(2004.2c)**

SOLUTION:

We assume that the outcome of each game is independent of the others.
Then we have a sequence of 8 Bernoulli trials where "success" means win for A.
If X denotes the total number of successes, then from standard result we know that $X \sim \text{Binomial}\left(8, \frac{2}{3}\right)$.
So the required probability is

$$\begin{aligned} P(X\geq 6) &= P(X=6)+P(X=7)+P(X=8) \\ &= \binom{8}{6}\left(\tfrac{2}{3}\right)^6\left(\tfrac{1}{3}\right)^2 + \binom{8}{7}\left(\tfrac{2}{3}\right)^7\left(\tfrac{1}{3}\right) + \binom{8}{8}\left(\tfrac{2}{3}\right)^8 \\ &= \left(\tfrac{1}{3}\right)^8\left[\frac{8\times 7}{2}2^6 + 8\times 2^7 + 2^8\right] \\ &= \cdots \\ &\quad \text{মাঝের অংকটুকু তুমি করে নিও।} \\ &= \frac{1024}{2187}. \end{aligned}$$

■

Example 5: Two dice are tossed together a number of times in succession. Find the minimum number of throws to ensure that the probability of obtaining six on both the dice at least once is greater than 0.5.[5] **(2009.1e)**

SOLUTION: এখানে ভালো করে বুঝে নাও যে random experiment-টা কী। এক সাথে দুটো ছক্কা চালা হয়েছে, এবং আমাদের মাথা ব্যথা হল দুটোতেই ছয় পড়েছে কিনা তা নিয়ে। সুতরাং আমাদের কাছে random experiment-টার খালি দুটোই outcome, হয় দুটোতেই ছয় পড়েছে, নয়তো অন্ততঃ একটাকে ছয় পড়েনি।

> We consider the Bernoulli random experiment of tossing the two dice together, where
>
> - "success" = both dice show 6,
> - "failure" = not "success"

এবার $P(\text{success})$ বার করা যাক--

> Assuming the dice to be fair and independent, the probability that both of them show 6 is
>
> $$\begin{aligned} & P(\text{first die shows } 6 \text{ and second die shows } 6) \\ = \; & P(\text{first die shows } 6) P(\text{second die shows } 6) \\ = \; & \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}. \end{aligned}$$
>
> If we repeat this random experiment n times independently, then let $X =$ number of successes.
> By standard result, $X \sim \text{Binomial}\left(n, \frac{1}{36}\right)$.
> So the chance of obtaining six on both the dice at least once is
>
> $$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - \binom{n}{0} \left(\tfrac{1}{36}\right)^{)} \left(1 - \tfrac{1}{36}\right)^{n} \\ &= 1 - \binom{n}{0} \left(\tfrac{35}{36}\right)^{n}. \end{aligned}$$

এবার দেখতে হবে সবচেয়ে ছোটো কোন n-এর জন্য এটা > 0.5 হয়--

> We want the minimum n such that
>
> $$\begin{aligned} & 1 - \left(\tfrac{35}{36}\right)^{n} > 0.5 \\ \iff \quad & \left(\tfrac{35}{36}\right)^{n} < 1 - 0.5 = 0.5 \end{aligned}$$
>
> '$>$'-টা উল্টে '$<$' হয়ে গেছে, লক্ষ করেছো নিশ্চয়ই?
> এখানে n-টা ওপরতলায় চলে গেছে, নামানোর জন্য log নিতে হবে।
>
> $$\begin{aligned} \iff \quad & n \log\left(\tfrac{35}{36}\right) < \log(0.5) \\ \iff \quad & n \log\left(\tfrac{36}{35}\right) > \log(2) \end{aligned}$$
>
> এই ধাপটা সাবধান! মনে রেখো যে $\log(1/a) = -\log(a)$.
> ওই মাইনাসটার জন্য inequality-টা আরেকবার ওল্টালো।
>
> $$\iff \quad n > \frac{\log(2)}{\log\left(\frac{36}{35}\right)} \approx 24.6.$$

> So the required answer is $n = 25$.

এই শেষ ধাপটা পেলাম, কারণ n একটা integer. ■

এইবার যে অংকটা করব সেটা একটু জটিল। সুখের কথা এই যে, এই অংকটা এ বইতে পরে কোথাও লাগবে না। পরীক্ষার দুশ্চিন্তা না থাকলে স্বচ্ছন্দে বাদ দিয়ে যেতে পারো।

Example 6: A and B play a game which must be either won or lost. If the probability that A wins any game is p, show that the probability that A wins m games before B wins n games is

$$\sum_{i=m}^{m+n-1} \binom{m+n-1}{i} p^i (1-p)^{m+n-1-i} = \int_0^p x^{m-1}(1-x)^{n-1}dx \Big/ \int_0^1 x^{m-1}(1-x)^{n-1}dx.$$

[7] **(2004.2a)**

SOLUTION: একটা উদাহরণ নিলে বুঝতে সুবিধা হবে। A আর B-এর মধ্যে খেলা হচ্ছে, প্রতি খেলাতেই হয় A জিতবে নয়তো B (মানে ড্রয়ের প্রশ্ন নেই)। ধরো $m = 3$ আর $n = 2$ নিলাম। তাহলে আমাদের event-টা হল "B দুবার জেতার আগেই A তিনবার জিতবে"। যেমন, ধরো যদি এরকম হয়--$ABAA$ (মানে প্রথম খেলায় A জিতেছে, দ্বিতীয় খেলায় B, এইরকম) তবে আমাদের event-টা ঘটেছে, কারণ প্রথম চারটে খেলার মধ্যে A তিনবার জিতে গেছে, কিন্তু B তখনও দুবার জিতে উঠতে পারে নি। কিন্তু যদি এরকম হত--ABB, তবে আমাদের আর আশা নেই, কারণ B দুবার জিতে নিয়েছে, A এখনও তিনবার জিতে উঠতে পারে নি। একটু চিন্তা করলেই বুঝবে যে, এখানে event-টা হতে পারে কেবলমাত্র এই কয়ভাবে--

$$AAA,\ BAAA,\ ABAA,\ AABA.$$

আমাদের কাজ হল এই event-টার probability বার করা (যেকোনো m, n-এর জন্য)। দুটো ফর্মুলা দেওয়া আছে, একটা sum দিয়ে, অন্যটা integral দিয়ে। প্রথমে sum-টা প্রমাণ করি। Sum-টার দিকে তাকালেই বুঝবে যে, এটা Binomial$(m+n-1, p)$-এর কিছু probability-র যোগফল। এখানে $m+n-1$ আসছে কোথা থেকে? সেটা বোঝার চেষ্টা করা যাক।

এর জন্য আমাদের উদাহরণটার দিকে আরেকবার তাকাও। লক্ষ কর যে, $m = 3, n = 2$ নিলে পুরো নিষ্পত্তি হয়ে যাচ্ছে প্রথম চারটে খেলার মধ্যেই। এর কারণ হল $m + n = 5$. সুতরাং $5 - 1$-এর মধ্যেই হয় A তিনটে জিতে নিয়েছে, নয়তো B দুটো জিতে নিয়েছে (কারণ A-র জয়ের সংখ্যা < 3 এবং B-এর জয়ের সংখ্যা < 2 হলে মোট খেলার সংখ্যা $\leq (3-1) + (2-1) = 3$ হত, মানে < 4 হত)। সুতরাং আমাদের event-টাকে আমরা চাইলে এভাবেও বলতে পারতাম--"প্রথম চারটে খেলার মধ্যে A অন্ততঃ তিনটে জিতেছে।" প্রশ্নে যেভাবে event-টা বর্ণনা করেছে তার থেকে এই বর্ণনাটা সহজ, কারণ এখানে খালি A-র জিতের হিসেব রাখলেই চলছে।

একইভাবে যে কোনো m, n-এর জন্যই প্রথম $m + n - 1$-টা খেলার মধ্যেই ব্যাপারটা নিষ্পত্তি হয়ে যেতে বাধ্য, এবং event-টাকে আমরা সহজতর ভাষায় বর্ণনা করতে পারি--"প্রথম $m + n - 1$-টা খেলার মধ্যে A অন্ততঃ m-টা জিতেছে।"

> **Proof of the summation formula:**
>
> Let M and N be the following two events:
>
> $M = \{A$ wins m games before B wins n games$\}$,
> $N = \{A$ wins at least m games out of the first before $m+n-1$ wins$\}$.

এখানে M হল প্রশ্নে দেওয়া বর্ণনাটা, আর N হল সহজতর বর্ণনাটা। প্রথমে প্রমাণ করে নিতে হবে, যে দুটোই আসলে একই event.

> Shall show that $M = N$.
>
> Step 1: Shall show $M \subseteq N$.
>
> Let, if possible, M occur, but not N.
>
> Since N has not occured, so out of the first $m+n-1$ games A has won $< m$ times. So B has won $> m+n-1-m = n-1$ times, *i.e.*, $\geq n$ times.

তার মানে B ইতিমধ্যেই অন্ততঃ n-টা জয় পেয়ে গেছে, কিন্তু A এখনও m-টা জয় পেয়ে উঠতে পারে নি।

> Thus, B has already won at least n games, but A is yet to win m games.
>
> So M cannot occur $(\Rightarrow\Leftarrow)$.
>
> So $M \subseteq N$.
>
> Step 2: Shall show $N \subseteq M$.
>
> Let N occur. Then A has already won $\geq m$ games within the first $m+n-1$ games.
>
> $\therefore B$ has won $\leq m+n-1-m = n-1$ games, *i.e.*, $< n$ games.
>
> So M has occured.
>
> So $M = N$.

এইবার Binomial-এর প্রবেশ--

> Let $X =$ number of wins for A in the first $m+n-1$ games.
>
> Then by standard result $X \sim \text{Binomial}(m+n-1, p)$.
>
> $$\therefore P(M) = P(N) = P(X \geq m) = \sum_{i=m}^{m+n-1} \binom{m+n-1}{i} p^i (1-p)^{m+n-1-i}.$$

এবার integral-ওয়ালা ফর্মুলাটা প্রমাণ করি--

> Proof of the integral formula:

এখানে একটা ছোটো ভুল আছে। Summation-এর ফর্মুলাটা যেকোনো m, n-এর জন্যই খাটে, $m = 0$ বা $n = 0$ হলেও আপত্তি নেই। কিন্তু $m = 0$ বা $n = 0$ হলে integral ফর্মুলাটা কিন্তু কাজ করে না। তাই আমরা এখানে ধরে নেব যে $m, n \geq 1$.

> We shall assume $m, n \in \mathbb{N}$ since the integrals in the RHS do not exist if $m = 0$ or $n = 0$.

Integral-গুলো দেখতে খুব একটা ভদ্র নয়। তবে মোটামুটি beta integral-এর মত, কখনো ওপরে আছে 1, কখনো p. আমরা এই general চেহারাটার একটা নাম দিয়ে নিই, তবে উত্তরটা গুছিয়ে লিখতে সুবিধা হবে।

Let $I_{m,n}(p) = \int_0^p x^{m-1}(1-x)^{n-1}dx$

আমাদের পরিকল্পনা হল বার বার করে integration by parts করা। প্রত্যেকবার by parts লাগালে খানিকটা করে অংশ integral-এর বাইরে চলে আসবে। সেইগুলো জমিয়ে জমিয়ে sum-এর ফর্মুলাটা পাব। আর শেষ পর্যন্ত যেটুকু integral পড়ে থাকবে সেটাকে কষে ফেলব।

$$\begin{aligned}
I_{m,n}(p) &= \int_0^p x^{m-1}(1-x)^{n-1}dx \\
&= \frac{1}{m}\int_{x=0}^p (1-x)^{n-1}d(x^m) \\
&= \frac{1}{m}x^m(1-x)^{n-1}\Big|_0^p - \frac{1}{m}\int_{x=0}^p x^m d\big((1-x)^{n-1}\big) \\
&= \frac{1}{m}p^m(1-p)^{n-1} + \frac{n-1}{m}\int_{x=0}^p x^m(1-x)^{n-2}dx \\
&= \frac{1}{m}p^m(1-p)^{n-1} + \frac{n-1}{m}I_{m+1,n-1}(p)
\end{aligned}$$

Thus

$$I_{m,n}(p) = \frac{1}{m}p^m(1-p)^{n-1} + \frac{n-1}{m}I_{m+1,n-1}(p) \qquad (*)$$

Putting $p=1$, we get

$$I_{m,n}(1) = \frac{n-1}{m}I_{m+1,n-1}(1). \qquad (**)$$

বুঝতেই পারছ যে, এই ফর্মুলাটা যতই লাগাবে, তাই m-টা এক এক করে বাড়তে থাকবে, আর n-টা কমতে থাকবে। সুতরাং শেষমেশ n-এর জায়গায় খালি 1 পড়ে থাকবে। সেই কেসটা কষে ফেলা সহজ--

Now for any k, we have $I_{k,1}(p) = \int_0^p x^{k-1}dx = \frac{p^k}{k}$.

এইবার তবে বারবার করে (**) লাগানো শুরু করি--

So, by using (**) repeatedly,

$$\begin{aligned}
I_{m,n}(1) &= \frac{(n-1)!}{m(m+1)\cdots(m+n-2)}I_{m+n-1,1} \\
&= \frac{(n-1)!}{m(m+1)\cdots(m+n-2)(m+n-1)} \\
&= \frac{(m-1)!(n-1)!}{(m+n-1)!}.
\end{aligned}$$

প্রশ্নটার দিকে আরেকবার তাকাও। আমাদের যে integral-এর ফর্মুলাটা প্রমাণ করতে দিয়েছে, সেখানে একটা integral-কে আরেকটা দিয়ে ভাগ করা আছে। এবার সেই ভাগটা করি--

Let $J_{m,n} = \frac{I_{m,n}(p)}{I_{m,n}(1)}$.

Then from (*) we have

$$J_{m,n} = \frac{1}{mI_{m,n}(1)} p^m (1-p)^{n-1} + \frac{n-1}{mI_{m,n}(1)} I_{m+1,n-1}(p)$$

But $\frac{1}{mI_{m,n}(1)} = \frac{1}{m} \times \frac{(m+n-1)!}{(m-1)!(n-1)!} = \binom{m+n-1}{m}$

and $\frac{n-1}{mI_{m,n}(1)} = \frac{n-1}{m} \times \frac{(m+n-1)!}{(m-1)!(n-1)!} = \frac{1}{I_{m+1,n-1}(1)}$.

So

$$J_{m,n} = \binom{m+n-1}{m} p^m (1-p)^{n-1} + \frac{I_{m+1,n-1}(p)}{I_{m+1,n-1}(1)} = \binom{m+n-1}{m} p^m (1-p)^{n-1} + J_{m+1,n-1}(p).$$

Applying this repeatedly,

$$J_{m,n} = \sum_{i=m}^{m+n-2} \binom{m+n-1}{i} p^i (1-p)^{m+n-1-i} + J_{m+n-1,1}(p) = \sum_{i=m}^{m+n-1} \binom{m+n-1}{i} p^i (1-p)^{m+n-1-i},$$

as required.

■

DAY 16
Standard discrete distribution families (part 2)

16.1 Binomial distribution (contd)

Binomial distribution-এর p.m.f.-এর সঙ্গে যে binomial theorem-এর ঘনিষ্ঠ যোগ আছে। এবার সেই যোগটা কাজে লাগবে expectation, variance, characteristic function ইত্যাদি বার করার সময়ে। binomial theorem-এর নানা রকম চেহারা হয়। তার মধ্যে যেটা আমাদের কাজে লাগবে সেটা চট্ করে মনে করে নিই।

Binomial theorem

For any two numbers, a, b, and any $n \in \mathbb{N}$ we have

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

ডানদিকের summation-টার চেহারা ভালো করে মনের মধ্যে এঁকে নাও--

- a-র exponent[1] আর b-এর exponent যোগ করলে সব সময়েই n হচ্ছে।
- ঠিক সেই একই n আছে $\binom{n}{k}$-এর উপর তলায়।
- sum-টা শুরু হচ্ছে $k = 0$ থেকে, শেষ হচ্ছে $k = n$-এ গিয়ে।

এবার আরও দুটো জিনিস মনে করিয়ে দিই, যেগুলো হায়ার সেকণ্ডারীতে শিখেছিলে--

- $k \geq 1$ হলে $k\binom{n}{k} = n\binom{n-1}{k-1}$ হয়।
- $k \geq 2$ হলে $k(k-1)\binom{n}{k} = n(n-1)\binom{n-2}{k-2}$.

প্রমাণ করা খুবই সোজা, সেটা তোমার জন্যেই রেখে দিলাম।
এবার তাহলে আমরা expectation আর variance বার করি।

> **THEOREM**
>
> If $X \sim \text{Binomial}(n, p)$, then
>
> $$E(X) = np \quad \text{and} \quad \text{var}(X) = np(1-p).$$

Proof: একটু আগে যা বললাম সেগুলো ব্যবহার করে এদের অনায়াসে প্রমাণ করতে পারবে। একটু ধরিয়ে দিই।

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} kP(X=k) \\ &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \end{aligned}$$

এবার $k\binom{n}{k} = n\binom{n-1}{k-1}$ লাগাতে চাই।
কিন্তু তার জন্য $k \geq 1$ চাই। লক্ষ কর যে, $k = 0$ হলে
সেই term-টা এমনিতেই শূন্য। তাই ওটাকে বাদ দিয়ে $k = 1$ থেকে শুরু করি--

$$\begin{aligned} &= \sum_{k=1}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\ &= \sum_{k=1}^{n} n\binom{n-1}{k-1}p^k(1-p)^{n-k} \quad \left[\, k\binom{n}{k} = n\binom{n-1}{k-1}\right] \\ &= \sum_{r=0}^{n-1} n\binom{n-1}{r}p^{r+1}(1-p)^{n-r-1} \quad \left[\text{putting } r = k-1\right] \\ &= np\sum_{r=0}^{n-1} \binom{n-1}{r}p^r(1-p)^{n-1-r} \end{aligned}$$

এই যে, np-টাকে বার করে নিলাম, এতে binomial theorem-এর চেহারাটা পাওয়া গেল।

$$= np[p + (1-p)]^{n-1} = np,$$

[1] একটা power-এ যেটা মাথায় থাকে, তাকে বলে exponent, যেমন a^2-এ exponent হল 2.

ঠিক যেমনটা দরকার ছিল।
$\text{var}(X)$ বার করার দায়িত্বটা তোমাকেই দিলাম। এখানে শুরু করতে হবে $E(X(X-1))$ বার করা দিয়ে। সেটা ঠিক $E(X)$-এর কায়দাতেই বেরোবে। তার পর তা থেকে বার করো $E(X^2) = E(X(X-1)) + E(X)$. এবার $\text{var}(X) = E(X^2) - \big(E(X)\big)^2$ বার করতে আর কতক্ষণ! **[Q.E.D]**

Example 7: Find the mean and variance of a binomial$\big(6, \frac{1}{3}\big)$ distribution.[1+1] **(2014.2a)**

SOLUTION:

> We know that if $X \sim \text{Binomial}(n,p)$ then
>
> $$\begin{aligned} E(X) &= np, \\ \text{var}(X) &= np(1-p). \end{aligned}$$
>
> Here $n = 6$ and $p = \frac{1}{3}$. So mean is $np = 2$, variance is $np(1-p) = \frac{4}{3}$.

■

এবার বলি binomial distribution-এর characteristic function কী। চাইলে সেখান থেকেও $E(X)$ আর $\text{var}(X)$ বার করা যেত।

THEOREM

The Binomial(n,p) distribution has characteristic function

$$Q(t) = \big(pe^{it} + (1-p)\big)^n \quad t \in \mathbb{R}.$$

এর প্রমাণ অতি সহজ। ধরো X-এর distribution হল Binomial(n,p). তাহলে

$$Q(t) = E(e^{itX}) = \sum_{x=0}^{n} \binom{n}{x} e^{itx} p^x (1-p)^{n-x} = \sum_{x=0}^{n} \binom{n}{x} (e^{it}p)^x (1-p)^{n-x} = \big(pe^{it} + (1-p)\big)^n.$$

Exercise 2: একই কায়দায় Binomial(n,p)-র MGF বার করো তো। ■

Exercise 3: Binomial(n,p)-র characteristic function ব্যবহার করে expectation আর variance বার করো। ■

Binomial(n,p) distribution-এ দুটো parameter আছে, n আর p. লক্ষ করো যে, যদি expectation আর variance জানা থাকে, তবে তা থেকে n আর p দুটোই বার করে দেওয়া যায়। কারণ variance হল $np(1-p)$, তাকে expectation (মানে np) দিয়ে ভাগ করলে $1-p$ বেরিয়ে যাচ্ছে। সেখানে থেকে p বার করে np-র মধ্যে বসিয়ে দিলে n-ও বেরিয়ে যাবে। নীচের অংকটা এই ব্যাপারটারই একটা প্রয়োগ।

Example 8: The mean and variance of a binomial distribution is μ and $k\mu$, respectively ($k > 0$ a constant). Find the probability of no success, one success and at least two successes.[5] **(2010.2c)**

SOLUTION:

> Let $X \sim \text{Binomial}(n, p)$.

এখানে আমরা মনে মনে X-কে কল্পনা করছি এইভাবে--একটা কয়েনকে n-বার টস করা হচ্ছে, যেখানে head পড়াটা হল "success". আমাদের X হল মোট success-এর সংখ্যা।

> Then
>
> $$\begin{aligned} E(X) &= np = \mu \quad (\text{given}) \\ \text{var}(X) &= np(1-p) = k\mu \quad (\text{given}). \end{aligned}$$

এবার n আর p বার করব--

> So $k = \frac{k\mu}{\mu} = 1 - p$, or $p = 1 - k$. Hence $n = \mu/p = \mu/(1-k)$.

একবার যেই n, p বেরিয়ে গেল, এবার যেকোনো probability-ই বার করা যাবে--

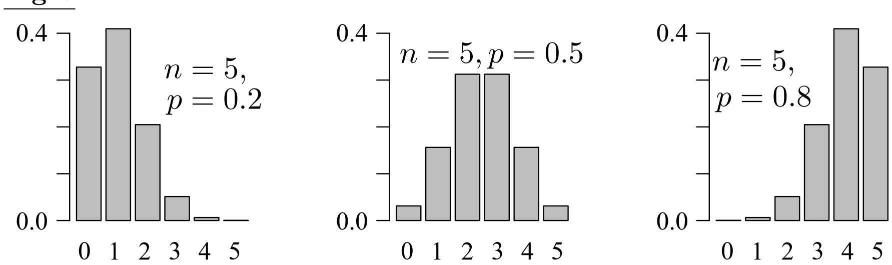
> Now
>
> $$\begin{aligned} P(\text{no success}) &= P(X=0) = \binom{n}{0} p^0 (1-p)^n = k^{\frac{\mu}{(1-k)}}, \\ P(1 \text{ success}) &= P(X=1) = \binom{n}{1} p^1 (1-p)^{n-1} = np(1-p)^{n-1} = \mu k^{\frac{\mu}{(1-k)}-1}, \\ P(\text{at least } 2 \text{ successes}) &= P(X \geq 2) = 1 - P(X=0) - P(X=1) \end{aligned}$$
>
> এখানে $P(X=2) + P(X=3) + \cdots + P(X=n)$। বার করতে হবে। অতগুলো যোগ না করে আমরা complement-টা নিয়ে কাজ করেছি।
>
> $$= 1 - k^{\frac{\mu}{(1-k)}} - \mu k^{\frac{\mu}{(1-k)}-1}.$$

■

Example 9: Show that the most probable number of successes in a Bernoullian sequence of n trials is the integer(s) i_m determined by the inequalities

$$(n+1)p - 1 \leq i_m \leq (n+1)p.$$

Fig 3



[5] **(2012.2b)**

SOLUTION: এখানে "number of successes in a Bernoullian sequence" নিয়ে কাজ হচ্ছে। তাই অংকটা আসলে binomial distribution-এর। প্রশ্নটা কী চাইছে সেটা বোঝার জন্য Fig 3-র দিকে তাকাও। এই ছবিটা আমরা আগেই দেখেছি। এখানে কিছু Binomial(n,p)-এর PMF আছে। প্রথম ছবিটায় একটা চূড়া আছে 1-এ, মানে $X \sim \text{Binomial}(5, 0.2)$ হলে X-এর পক্ষে 1 হওয়ার probability সবচেয়ে বেশী, বা ঘুরিয়ে বললে 1 হল Binomial$(5, 0.2)$-এর most probable value. এই "most probable value" জিনিসটার আরেকটা নাম হল **mode.** একইভাবে তৃতীয় ছবিটা বলছে যে Binomial$(5, 0.8)$-এর mode হল 4. মাঝের ছবিটায় কিন্তু দুটো চূড়া আছে পাশাপাশি 2 আর 3-এ। যদি n আর p দেওয়া থাকে তবে কোথায় কোথায় চূড়া হবে সেটা বার করার একটা ফর্মুলা প্রমাণ করতে বলেছে এই অংকটায়। ফর্মুলাটা সহজই, $[(n+1)p-1, (n+1)p]$-এর মধ্যে যেসব integer আছে সেখানেই চূড়া থাকবে। যেমন মাঝের ছবিটার বেলায় $n=5$ আর $p=0.5$. তাই $(n+1)p=3$. সুতরাং দেখতে হবে $[3-1,3]=[2,3]$-এর মধ্যে কটা integer আছে। ঠিক দুটোই আছে, 2 আর 3. এবং ছবিতে সত্যিই দেখা যাচ্ছে যে, ঠিক ওই দুটো জায়গাতেই চূড়া আছে।

এবার প্রমাণে হাত দেওয়া যাক। লক্ষ কর যে, তিনটে ছবিতেই প্যাটার্নটা একই--প্রথমে চূড়া অব্ধি উঠছে, তারপর নামছে। উঠছে মানে $P(X=i) < P(X=i+1)$, আর নামছে মানে $P(X=i) > P(X=i+1)$. যদি দুটো চূড়ো থাকে তবে তাদের উচ্চতা সমান, মানে $P(X=i) = P(X=i+1)$. আমরা এই তিনটে কেসকেই একসঙ্গে $\lesseqgtr$ লিখে করব।

> Let $X \sim \text{Binomial}(n,p)$.
> Then for $i = 0, ..., n-1$,
>
> $$\begin{aligned} & P(X=i) \lesseqgtr P(X=i+1) \\ \iff & \frac{P(X=i)}{P(X=i+1)} \lesseqgtr 1 \\ \iff & \frac{\binom{n}{i} p^i (1-p)^{n-i}}{\binom{n}{i+1} p^{i+1} (1-p)^{n-i-1}} \lesseqgtr 1 \\ \iff & \cdots \end{aligned}$$
>
> এখানে $\binom{n}{x}$-এর সংজ্ঞা লাগিয়ে দিলেই কিছু কাটাকাটি হবে।
>
> $$\begin{aligned} \iff & (i+1)(1-p) \lesseqgtr (n-i)p \\ \iff & \cdots \end{aligned}$$
>
> আবার একটা মামূলী ধাপ। তাই লিখে দেখাচ্ছি না।
>
> $$\iff \quad i+1 \lesseqgtr (n+1)p.$$
>
> Thus, as i increases, the value of $P(X=i)$ first increases and then decreases.

অর্থাৎ আমরা ছবি থেকে যে প্যাটার্ণটা দেখেছিলাম (মানে প্রথমে উঠে তারপর নামা) সেটা প্রমাণ হল। এবার চূড়োটা কোথায় দেখব। যদি i_m-এ চূড়া থাকে, তবে $P(X=i_m)$-এর চেয়ে বড় probability হতে পারে না, সুতরাং--

> We need $P(X=i_m) \geq P(X=i_m+1)$ and so
>
> $$i_m + 1 \geq (n+1)p.$$
>
> Also we need $P(X=i_m-1) \leq P(X=i_m)$ and so
>
> $$i_m \leq (n+1)p.$$

> Combining, we have
> $$(n+1)p - 1 \le i_m \le (n+1)p,$$
> as required.

■

Example 10: In a Bernoullian sequence of n trials with constant probability of success p, find the most probable number of successes.[5] **(2010.2a)**
SOLUTION: আগের অংকটাই। ■

একটু আগেই বললাম যে, most probable value-র আরেকটা নাম হল mode. সেই ভাষাটা ব্যবহার করা হয়েছে নীচের অংকটায়।

Example 11: Find the mode of a binomial$(6, \frac{1}{3})$ distribution.[1] **(2014.2a)**

SOLUTION:

> We know that for Binomial(n,p) distribution, mode is any integer m satisfying
> $$(n+1)p - 1 \le m \le (n+1)p.$$

এইটাই আমরা আগের অংকে প্রমাণ করেছি। এটা একটা standard result, তাই এখানে ব্যবহার করা যায়।

> Here $n = 6$ and $p = \frac{1}{3}$. So mode is any integer m satisfying
> $$\frac{7}{3} - 1 \le m \le \frac{7}{3},$$
> or $m = 2$.

■

Example 12: Prove that the sum of two binomially distributed independent random variables with parameters (n_1, p) and (n_2, p) is again a binomially distributed random variable with parameters $(n_1 + n_2, p)$.[5] **(2013.4b)**
SOLUTION: এই অংকটা আমরা তিনভাবে করব। প্রথমটা একেবারেই সোজা, কোনো অংকই নেই বলতে গেলে। খালি কয়েন টসের গল্পটা ব্যবহার করব--

> Let $X_1 \sim \text{Binomial}(n_1, p)$ and $X_2 \sim \text{Binomial}(n_2, p)$ be two independent random variables.
>
> We know that a Binomial(n,p) random variable can be considered as the total number of heads obtained in n independent tosses of a coin with $P(\text{head}) = p$.
>
> So behind each X_i we have n_i independent tosses of the same coin.

So there are n_1+n_2 tosses of the coin in total. Since the X_i's are independent, so all these tosses are independent.

Also X_1+X_2 is the total number of heads out of all these n_1+n_2 independent tosses of this coin with $P(head)=p$. So $X_1+X_2 \sim \text{Binomial}(n_1+n_2,p)$ distribution.

যদি এত সংক্ষিপ্ত উত্তরে তোমার মন না ভরে, তবে PMF ধরে ধরে অংক কষেও প্রমাণ করা যায়--

<u>বিকল্প পদ্ধতি</u>

Let $X_1 \sim \text{Binomial}(n_1,p)$ and $X_2 \sim \text{Binomial}(n_2,p)$ be two independent random variables.

Then X_1 can take any value in $\{0,1,...,n_1\}$ and X_2 can take any value is $\{0,1,...,n_2\}$.

$\because X_1,X_2$ are independent, X_1+X_2 can take any values in $\{0,1,...,n_1+n_2\}$.

Shall show

$$\forall k\in\{0,1,...,n_1+n_2\} \quad P(X_1+X_2=k)=\binom{n_1+n_2}{k}p^k(1-p)^{n_1+n_2-k}.$$

Take any $k\in\{0,1,...,n_1+n_2\}$.

$$\begin{aligned}
P(X_1+X_2=k) &= \sum_{i=0}^{n_1} P(X_1=i,X_2=k-i)\\
&\qquad \text{কারণ } X_1+X_2=k \text{ বলে, ওদের একজন } i \text{ হলে}\\
&\qquad \text{অন্যজন } k-i \text{ হতে বাধ্য।}\\
&= \sum_{i=0}^{n_1} P(X_1=i)P(X_2=k-i) \quad [\because \text{independent}]\\
&= \sum_{i=0}^{n_1}\binom{n_1}{i}p^i(1-p)^{n_1-i}\binom{n_2}{k-i}p^{k-i}(1-p)^{n_2-(k-i)}\\
&= \sum_{i=0}^{n_1}\binom{n_1}{i}\binom{n_2}{k-i}p^k(1-p)^{n_1+n_2-k}\\
&= p^k(1-p)^{n_1+n_2-k}\sum_{i=0}^{n_1}\binom{n_1}{i}\binom{n_2}{k-i}\\
&= p^k(1-p)^{n_1+n_2-k}\binom{n_1+n_2}{k}
\end{aligned}$$

by standard result from binomial theorem, as required.

বেশ লম্বা প্রমাণ! যদি characteristic function ব্যবহার করি, তবে সংক্ষেপে হবে--

<u>বিকল্প পদ্ধতি</u>

> Let $X_1 \sim \text{Binomial}(n_1, p)$ and $X_2 \sim \text{Binomial}(n_2, p)$ be two independent random variables.
>
> Then, from standard result, X_k has c.f.
>
> $$Q_k(t) = (pe^{it} + 1 - p)^{n_k}.$$

এখানে আমরা X-এর নীচের k ব্যবহার করেছি, কারণ i ব্যবহার করলে complex সংখ্যা i-এর সঙ্গে গুলিয়ে যাবে!

> Since X_1, X_2 are independent, hence the c.f. of $X_1 + X_2$ is
>
> $$Q(t) = Q_1(t) \times Q_2(t) = (pe^{it} + 1 - p)^{n_1+n_2},$$
>
> which we recognise as the c.f. of the $Binomial(mn, p)$ distribution.
>
> Hence, by characterising property of characteristic function, the distribution of $X_1 + \cdots + X_m$ is $Binomial(mn, p)$.

■

Example 13: If X is a Binomial(n, p) variate, then prove that

$$\mu_{k+1} = p(1-p)\left(nk\mu_{k-1} + \frac{d\mu_k}{dp}\right),$$

where μ_k is the k-th order central moment.[5] **(2014,2007)**

SOLUTION:

> We know that $E(X) = np$.
>
> $$\therefore \mu_k = \sum_{x=0}^{n} (x - np)^k \binom{n}{x} p^x (1-p)^{n-x}.$$

এবার আমরা এটাকে p-এর সাপেক্ষে differentiate করব। লক্ষ করো যে, sum-এর মধ্যে তিনটে জিনিসের গুণফল আছে (এখানে $\binom{n}{x}$-টাকে হিসেবে ধরছি না, কারণ ওতে কোনো p নেই। তিনটে জিনিসের গুণফলকে differentiate করলে তিনটে অংশ হবে--

So

$$\begin{aligned}\frac{d\mu_k}{dp} &= \sum_{x=0}^{n}\binom{n}{x}\underbrace{k(x-np)^{k-1}(-n)}p^x(1-p)^{n-x}\\ &\quad+\sum_{x=0}^{n}\binom{n}{x}(x-np)^k\underbrace{xp^{x-1}}(1-p)^{n-x}\\ &\quad-\sum_{x=0}^{n}\binom{n}{x}(x-np)^k p^x\underbrace{(n-x)(1-p)^{n-x-1}}\end{aligned}$$

সাবধান, এখানে কিন্তু differentiation-টা হচ্ছে p-এর সাপেক্ষে, x-এর সাপেক্ষে নয়!
এবার আমরা প্রথম অংশটাকে একটু দলাইমলাই করি--

Now

$$\begin{aligned}&\sum_{x=0}^{n}\binom{n}{x}k(x-np)^{k-1}(-n)p^x(1-p)^{n-x}\\ &= -nk\underbrace{\sum_{x=0}^{n}\binom{n}{x}(x-np)^{k-1}p^x(1-p)^{n-x}}_{\mu_{k-1}}\\ &= -nk\mu_{k-1}.\end{aligned}$$

এবার যেটা প্রমাণ করতে বলেছিল, তার ডানদিকটা নিয়ে শুরু করি। সেটাকে অনেক লড়াই করে দেখাব μ_{k+1}-এর সমান। ভালো করে দম নিয়ে নাও।

So

$$
\begin{aligned}
& p(1-p)\left[\frac{d\mu_k}{dp} + nk\mu_{k-1}\right] \\
= \; & p(1-p)\left[\sum_{x=0}^{n}\binom{n}{x}(x-np)^k x p^{x-1}(1-p)^{n-x} - \sum_{x=0}^{n}\binom{n}{x}(x-np)^k p^x (n-x)(1-p)^{n-x-1}\right] \\
= \; & (1-p)\sum_{x=0}^{n} x\binom{n}{x}(x-np)^k p^x(1-p)^{n-x} - p\sum_{x=0}^{n}\binom{n}{x}(x-np)^k (n-x) p^x (1-p)^{n-x}
\end{aligned}
$$

এখন দুটো sum-এর শেষেই $p^x(1-p)^{n-x}$ এসে গেছে।
এবার দ্বিতীয় sum-এর $(n-x)$-এর ব্র্যাকেট তুলব।
ফলে ওই sum-টা ভেঙে দুটো sum হয়ে যাবে।

$$
\begin{aligned}
= \; & (1-p)\sum_{x=0}^{n} x\binom{n}{x}(x-np)^k p^x(1-p)^{n-x} \\
& -np\sum_{x=0}^{n}\binom{n}{x}(x-np)^k p^x(1-p)^{n-x} + p\sum_{x=0}^{n} x\binom{n}{x}(x-np)^k p^x(1-p)^{n-x}
\end{aligned}
$$

এবার প্রথম sum-টার আগে যে $(1-p)$-টা আছে, তার ব্র্যাকেট তুলব।
সুতরাং ওই sum-টা ভেঙে দুই টুকরো হবে।

$$
\begin{aligned}
= \; & \sum_{x=0}^{n} x\binom{n}{x}(x-np)^k p^x(1-p)^{n-x} - p\sum_{x=0}^{n} x\binom{n}{x}(x-np)^k p^x(1-p)^{n-x} \\
& -np\sum_{x=0}^{n}\binom{n}{x}(x-np)^k p^x(1-p)^{n-x} + p\sum_{x=0}^{n} x\binom{n}{x}(x-np)^k p^x(1-p)^{n-x}
\end{aligned}
$$

এবার দ্যাখো, দ্বিতীয় আর চতুর্থ sum-দুটো কেটে যাচ্ছে।

$$
= \sum_{x=0}^{n} x\binom{n}{x}(x-np)^k p^x(1-p)^{n-x} - np\sum_{x=0}^{n}\binom{n}{x}(x-np)^k p^x(1-p)^{n-x}
$$

দুটোকে জুড়ে একটা sum করে নাও--

$$
\begin{aligned}
= \; & \sum_{x=0}^{n}\binom{n}{x}(x-np)^{k+1} p^x(1-p)^{n-x} \\
= \; & \mu_{k+1},
\end{aligned}
$$

completing the proof.

■

DAY 17
Standard discrete distribution families (part 3)

17.1 Poisson distribution

মনে করো, আমি একটা কয়েন নিলাম যার head পড়ার probability খুব কম, ধরো 0.03, মানে 100 বার টস করলে গড়ে মোটে 3 বার head পাব। এই কয়েনটাকে এবার যদি 500 বার টস করা হয়, তবে বলতে পারবে ঠিক 50-টা head পড়ার probability কত? আমরা binomial distribution শিখেছি, তাই উত্তরটা লিখে দেওয়া কঠিন নয়--

$$\binom{500}{50} 0.03^{50} (1-0.03)^{500-50}.$$

এবার ধরো বললাম এটা বার করতে। খালি হাতে করা তো দুঃসাধ্য বটেই, এমন কি ক্যালকুলেটরেও কাজটা প্রায় অসম্ভব। ওই $\binom{500}{50}$-টা একটা বি-শা-ল সংখ্যা আর 0.03^{50}-টা একটা পুঁ-চ-কে সংখ্যা, এবং $(1-0.03)^{500-50}$-টাও বেশ ছোটোই। ওদের গুণ করে কী দাঁড়াবে সেটা বার করা সহজ নয় মোটেই। এইখানে একটা approximation খুব কাজে আসে। যদি $p = P(\text{head})$-টা খুব ছোটো হয়, আর টস করা হয় n-বার, যেখানে n খুব বড়, তাহলে এই জিনিসটা হয়--

$$\forall i \in \{0, 1, ..., n\} \quad \binom{n}{i} p^i (1-p)^{n-i} \approx \frac{e^{-np}(np)^i}{i!}.$$

ডানদিকে যেটা রয়েছে সেটা অনায়াসেই ক্যালকুলেটর দিয়ে বার করা যায়। অবশ্য এতে খালি একটা approximate উত্তর আসে। স্বভাবতঃই প্রশ্ন ওঠে যে, approximation-টা কীভাবে এল, এবং এর জন্য কোনো শর্ত লাগে কিনা। তার হদিশই রয়েছে নীচের অংকটায়।

Example 14: If $p = \frac{\mu}{n}$, where μ is a positive constant and $0 < p < 1$; n is a positive integer, then prove that

$$\lim_{n\to\infty} \binom{n}{i} p^i (1-p)^{n-i} = e^{-\mu} \frac{\mu^i}{i!}.$$

[5] **(2011.2a)**

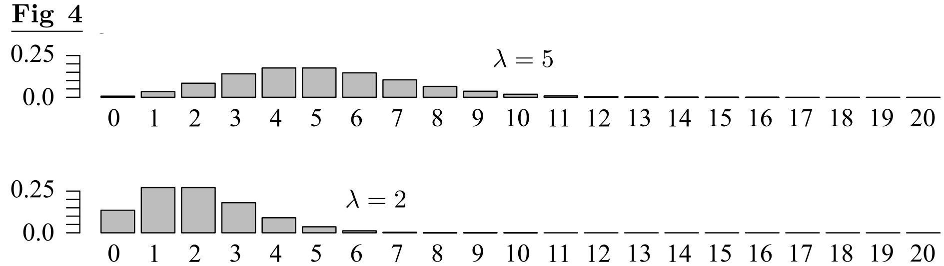
SOLUTION: এক্ষুণি যা বললাম তার সঙ্গে এই অংকটাকে মিলিয়ে নাও। আমরা বলেছিলাম যে n খুব বড় আর p খুব ছোটো। এই অংকে সেটাকেই আরও খুঁটিয়ে বলেছে যে, $p = \mu/n$, যেখানে $\mu(>0)$ কোনো একটা constant. ফলে n যত বাড়ছে, p ততই কমছে।

We have

$$\binom{n}{i} p^i (1-p)^{n-i} = \frac{n(n-1)\cdots(n-i+1)}{i!} \left(\frac{\mu}{n}\right)^i \left(1 - \frac{\mu}{n}\right)^{n-i}$$

এখানে আমরা যাবতীয় p-কে $\frac{\mu}{n}$ আকারে লিখেছি। আমাদের দেখাতে হবে যে limit-টা $e^{-\mu}\frac{\mu^i}{i!}$. এর মধ্যে $\mu^i/i!$ অংশটা কিন্তু ইতিমধ্যেই এসে গেছে! সেটাকে আলাদা করে সরিয়ে রাখো--

Fig 4



$$= \frac{\mu^i}{i!} \times \frac{n(n-1)\cdots(n-i+1)}{n^i}\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-i}$$

Now, as $n \to \infty$,

$$\begin{aligned}\frac{n(n-1)\cdots(n-i+1)}{n^i} &= \frac{n}{n} \times \frac{n-1}{n} \times \cdots \times \frac{n-i+1}{n} \to 1,\\ \left(1-\frac{\mu}{n}\right)^n &\to e^{-\mu},\\ \left(1-\frac{\mu}{n}\right)^{-i} &\to 1.\end{aligned}$$

এর মধ্যে তৃতীয় limit-টা সহজ, দ্বিতীয়টা standard. প্রথমটার জন্য মনে রেখো যে i কিন্তু এখানে স্থির আছে, সেটা মোটেই ∞-তে যাচ্ছে-টাচ্ছে না। যেমন $i = 2$ নিলে $\frac{n(n-1)}{n^2} \to 1$.

Hence

$$\lim_{n\to\infty}\binom{n}{i}p^i(1-p)^{n-i} = e^{-\mu}\frac{\mu^i}{i!},$$

as required.

■

এই approximation-টা এতই কাজ আসে যে এর একটা আলাদা নামই দেওয়া হয়েছে--Poisson distribution. এখানে Poisson-এর উচ্চারণ হল পোয়াস্‌অঁ।

Example 15: Define Poisson distribution of a random variable X. **(2009,2007)**

SOLUTION:

DEFINITION: **Poisson distribution**

Let X be a random variable taking only nonnegative integer values. It is said to have a Poisson distribution if it has PMF

$$P(X=n) = \begin{cases} \frac{e^{-\lambda}\lambda^n}{n!} & \text{if } n = 0,1,2,... \\ 0 & \text{otherwise} \end{cases}$$

Here $\lambda > 0$ is a parameter. We write $X \sim \text{Poisson}(\lambda)$.

■

Fig 4-এ বিভিন্ন Poisson distribution-এর PMF-এর ছবি দিয়েছি।
নীচের অংকটায় আমরা হাতে কলমে দেখতে পারব যে Poisson approximation to binomial-টা কত কাছাকাছি উত্তর দেয়।

Example 16: The probability of hitting a target is 0.002 for each shot. A shooter aims to hit

the target. Find the probability of the shooter's success with two or more shots, if the number of shots is 1000.[4] **(2004.2b)**

SOLUTION: এই অংকের গল্পটা বুঝে নাও। একটা লোক একটা টার্গেটে গুলি করছে। আনাড়ি লোক, তাই গুলিগুলো ফস্কাবার সম্ভাবনা খুবই বেশী। বলা আছে যে, টার্গেটে গুলি লাগার সম্ভাবনা হল 0.002, অর্থাৎ কিনা 1000 বার গুলি ছুঁড়লে গড়ে দুটো গুলি টার্গেটে গিয়ে লাগবে, বাকিগুলো ফস্কাবে। এবার ধরো লোকটা সত্যিই 1000 বার গুলি ছুঁড়েছে। তবে অন্ততঃ দুটো টার্গেটে লাগার probability বার করতে বলেছে।
এই অংকটা আমরা দুইভাবে করব। একবার binomial দিয়ে, আর অন্যবার Poisson approximation to binomial দিয়ে।

> Let X be the number of successful shots out of the 1000 shots.
> Assuming the shots to be independent and identically distributed,
> we have $X \sim \text{Binomial}(1000, 0.002)$.

এটা হল কারণ প্রতিটা গুলি ছোঁড়া যেন একটা কয়েন টসের মত, head মানে টার্গেটে লেগেছে, আর tail মানে লাগেনি।

> So the required probability is
>
> $$\begin{aligned} P(X \geq 2) &= 1 - P(X=0) - P(X=1) \\ &= 1 - \binom{1000}{0}(0.002)^0(0.998)^{1000} - \binom{1000}{1}(0.002)(0.998)^{999} \\ &= 1 - (0.998)^{1000} - 2(0.998)^{999} \\ &= \cdots = 0.5943. \end{aligned}$$

এইখানেই উত্তরটা শেষ। কিন্তু এবার তুলনার জন্য আমরা Poisson approximation to binomial লাগিয়ে দেখব। যেহেতু এখানে binomial দিয়ে সমাধানটা মোটেই জটিল কিছু ছিলনা, তাই approximation-টার কোনো দরকার নেই। তাও আমরা খালি তুলনা করে দেখার জন্য approximation-টা করব--

> Also, we know that if $n \to \infty$ with $np = \lambda > 0$, fixed, then $\text{Binomial}(n, p) \approx \text{Poisson}(\lambda)$.
> Here $\lambda = np = 1000 \times 0.002 = 2$. Assuming $n = 1000$ to be large enough for this approximation to be good,
>
> $$P(X \geq 2) = 1 - P(X=0) - P(X=1) \approx 1 - e^{-2}\frac{2^0}{0!} - e^{-2}\frac{2^1}{1!} = 1 - \frac{3}{e^2} = 0.5940.$$

কোনো approximation ছাড়া করে উত্তর এসেছিল 0.5943, আর এখানে এল 0.5940. সুতরাং approximation-টা যথেষ্ট কাছাকাছি গেছে! এখানে আমরা $P(X \geq 2)$ বার করছিলাম। যদি $P(X \geq 10)$ বার করতে দিত, তবে approximation ছাড়া এগোনো অনেক বেশী কঠিন হত। ■

এই যে approximation-টা আমরা শিখলাম এর ভিত্তিতে অনেক সময়ে বলা হয় যে--

"Number of occurrences of a rare event has a Poisson distribution."

যেহেতু Poisson distribution-এর বিভিন্ন প্রয়োগের পিছনেই এই ধারণাটা কাজ করে, তাই এর অর্থটা ভালো করে বুঝে রাখা যাক। আমরা জানি যে, কোনো Bernoulli random experiment-কে যদি independently করা যায় n-বার,

তবে success-এর সংখ্যা একটা random variable হয় যার distribution হবে Binomial(n, p). এখানে p হল $P(\text{success})$. যদি p খুব ছোটো হয়, তবে success খুবই বিরল ঘটনা, যাকে ইংরিজিতে বলে rare event. আমাদের random variable-টা হচ্ছে সেই rare event কতবার ঘটেছে তার গুণতি, সেই কারণেই "number of occurrences of a rare event" বলা। এরকম নানা উদাহরণ কল্পনা করা যায়--

- বইয়ের পাতায় ছাপার ভুলের সংখ্যা (অবশ্য একেকটা বইতে অ্যাত্তো ভুল থাকে যে তাকে আর rare event বলা যায় না!)
- ভদ্রসভ্য রাস্তায় দুর্ঘটনার সংখ্যা,
- আগেকার দিনে মোট phone call-এর সংখ্যাকেও Poisson ধরা হত, কারণ তখন phone ব্যাপারটা এত জনপ্রিয় ছিল না।

এরকম rare event-এর অংক আমরা একটু বাদেই দেখব। কিন্তু তার আগে Poisson distribution-এর expectation, variance ইত্যাদি বার করে নিই।

THEOREM

If $X \sim \text{Poisson}(\lambda)$, then

$$E(X) = \lambda \quad \text{and} \quad \text{var}(X) = \lambda.$$

Proof: প্রমাণটা খুবই সহজ। একটু ধরিয়ে দিচ্ছি, বাকিটা নিজে চেষ্টা করো। আমরা জানি যে, যে কোনো $a \in \mathbb{R}$-এর জন্যই

$$e^a = 1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \cdots .$$

ডানদিকে চেহারাটা খেয়াল রাখো। প্রত্যেকটা term-এ উপরতলায় যত power, নীচে ঠিক সেটারই factorial. যদি summation দিয়ে লিখি তবে প্যাটার্ণটা আরো ভালো করে ফুটে উঠবে--

$$e^a = \sum_{k=0}^{\infty} \frac{a^k}{k!}.$$

এবার $E(X)$-এর ফর্মুলাটা লেখা যাক--

$$E(X) = \sum_{k=0}^{\infty} kP(X = k) = \sum_{k=0}^{\infty} ke^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!}$$

এইবার আমরা k-র সঙ্গে $k!$ কাটাকাটি করে $(k-1)!$ করার চেষ্টা করব। কিন্তু সেটা $k = 0$-র জন্য খাটবে না। কিন্তু $k = 0$ হলে সেই term-টা আসলে 0, তাই নিশ্চিন্তে বাদ দেওয়া যায়--

$$= e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \sum_{r=0}^{\infty} \frac{\lambda^{r+1}}{r!} \quad \left[\text{putting } r = k - 1 \right]$$

$$= \lambda e^{-\lambda} \sum_{r=0}^{\infty} \frac{\lambda^r}{r!}.$$

এবার কী করতে হবে সেটা নিশ্চয়ই আর বলে দিতে হবে না?
$\text{var}(X)$ বার করার জন্য একইরকম কায়দা করতে হবে। আমরা জানি $\text{var}(X) = E(X^2) - \big(E(X)\big)^2$. এর মধ্যে $E(X)$ তো বার করলামই। বাকি আছে $E(X^2)$ বার করা। এর জন্য প্রথমে $E(X(X-1))$ বার করো summation ব্যবহার করে, ঠিক যেভাবে $E(X)$ বার করলাম।
তা থেকে $E(X^2)$ পেয়ে যাবে, কারণ $E(X^2) = E(X(X-1)) + E(X)$. **[Q.E.D]**

এবার একটা অংক দেখি যেখানে সেই rare event-এর গল্পটা কাজে দেবে।

Example 17: There are 500 misprints in a book of 500 pages. What is the probability that a given page will contain at least 3 misprints?[5] **(2009.2b)**

SOLUTION:

> Let X be the number of misprints in a randomly selected page.
>
> Assuming that a misprint is a rare event, we take $X \sim \text{Poisson}(\lambda)$ for some $\lambda > 0$.

Family-টা সনাক্ত করেছি, এবার family-র সদস্যটাকে সনাক্ত করতে হবে, মানে λ বার করার চেষ্টা করতে হবে।

> On an average each page contains $\frac{500}{500} = 1$ misprint. So $E(X) = 1$, ie, $\lambda = 1$.

যেই λ পেয়ে গেলাম, এবার যেকোনো probability বার করে ফেলা যাবে।

> Thus, the required probability is
>
> $$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) \\ &= 1 - P(X=0) - P(X=1) - P(X=2) \\ &= 1 - \frac{e^{-1}1^0}{0!} - \frac{e^{-1}1^1}{1!} - \frac{e^{-1}}{2!} \\ &= 1 - \frac{5}{2e}. \end{aligned}$$

■

এবার আমরা Poisson(λ)-র characteristic function আর moment generating function শিখব।

THEOREM

If $X \sim \text{Poisson}(\lambda)$, then its characteristic function is

$$Q(t) = \exp[\lambda(e^{it} - 1)], \quad t \in \mathbb{R},$$

and moment generating function is

$$\phi(t) = \exp[\lambda(e^t - 1)], \quad t \in \mathbb{R}.$$

প্রমাণ খুবই সহজ। যেমন ধরো characteristic function-এর বেলায়

$$Q(t) = E(e^{itX}) = \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!}.$$

এবার নিশ্চয়ই exponential series-এর চেনা চেহারাটা চোখে পড়ছে?

Example 18: Prove that the sum of two independent Poisson variates having parameters μ_1, μ_2 is a Poisson variate with parameter $\mu_1 + \mu_2$. **(2009,2007)**

SOLUTION: এই অংকটা দুভাবে করব। প্রথমে সরাসরি PMF ব্যবহার করে। পরে characteristic function দিয়ে।

Let $X \sim \text{Poisson}(\mu_1)$ and $Y \sim \text{Poisson}(\mu_2)$ and X, Y be independent.
Then X, Y both take nonnegative integer values, and since they are independent $X + Y$ can also take any nonnegative integer value.
Shall show that $X + Y \sim \text{Poisson}(\mu_1 + \mu_2)$, ie,

$$\forall n \in \{0, 1, 2, ...\} \quad P(X + Y = n) = \frac{e^{-(\mu_1+\mu_2)}(\mu_1 + \mu_2)^n}{n!}.$$

Take any $n \in \{0, 1, 2, ...\}$.
Then

$$P(X + Y = n) = \sum_{k=0}^{n} P(X = k, Y = n - k)$$

যেহেতু $X + Y = n$, তাই X অবশ্যই 0 থেকে k-র মধ্যে কিছু একটা হবে। সেটাকে যদি k বলি, তবে $Y = n - k$ হতে বাধ্য।

$$= \sum_{k=0}^{n} P(X = k)P(Y = n - k) \quad [\because \text{independent}]$$

$$= \sum_{k=0}^{n} \frac{e^{-\mu_1}\mu_1^k}{k!} \frac{e^{-\mu_2}\mu_2^{n-k}}{(n-k)!} \quad [\because X \sim \text{Poisson}(\mu_1), \quad Y \sim \text{Poisson}(\mu_2)]$$

$$= e^{-(\mu_1+\mu_2)} \sum_{k=0}^{n} \frac{\mu_1^k}{k!} \frac{\mu_2^{n-k}}{(n-k)!}$$

$$= \frac{e^{-(\mu_1+\mu_2)}}{n!} \sum_{k=0}^{n} \binom{n}{k} \mu_1^k \mu_2^{n-k}$$

$$= \frac{e^{-(\mu_1+\mu_2)}}{n!} (\mu_1 + \mu_2)^n \quad [\text{by binomial theorem}],$$

as required.

এবার একই অংক করব characteristic function দিয়ে।

বিকল্প পদ্ধতি

> Let $X \sim \text{Poisson}(\mu_1)$ and $Y \sim \text{Poisson}(\mu_2)$ and X, Y be independent.
> Then characteristic functions of X, Y are
>
> $$Q_X(t) = \exp[\mu_1(e^{it}-1)] \text{ and } Q_Y(t) = \exp[\mu_2(e^{it}-1)] \text{ for } t \in \mathbb{R}.$$
>
> Since X, Y are independent, hence the characteristic function of $X+Y$ is
>
> $$Q(t) = Q_X(t)Q_Y(t) = \exp[\mu_1(e^{it}-1)] \times \exp[\mu_2(e^{it}-1)] = \exp[(\mu_1+\mu_2)(e^{it}-1)],$$
>
> which is the characteristic function of $\text{Poisson}(\mu_1+\mu_2)$ distribution.
> Hence, by the characterising property of characteristic functions, $X+Y \sim \text{Poisson}(\mu_1+\mu_2)$, as required.

■

Example 19: If X, Y are independent Poisson variates with parameters μ_1, μ_2, respectively, then show that $X+Y$ is a Poisson variate with parameter $\mu_1+\mu_2$.[5] **(2011.2c)**
SOLUTION: আগের অংকেই করেছি। ■

এইবার দুটো অংক করব যেগুলো একটু জটিল লাগতে পারে। ভয় নেই, এগুলো পরে কোথাও লাগবে না। তাই প্রথমবার পড়ার সময়ে চাইলে ওদের বাদ দিয়ে যেতে পারো।

Example 20: If X is a Poisson variate with parameter μ, show that

$$P(X \le n) = \frac{1}{n!}\int_{\mu}^{\infty} e^{-x}x^n dx,$$

where n is any positive integer.[5] **(2011,2008)**
SOLUTION: এই অংকটা খালি integration by parts-এর খেলা। আমরা $\int_{\mu}^{\infty} e^{-x}x^n dx$-এর উপর বারবার integration by parts লাগিয়ে ওটাকে সহজ করে ফেলতে পারব।

$$\begin{aligned}
I_n := \int_{\mu}^{\infty} e^{-x}x^n dx &= -\int_{x=\mu}^{\infty} x^n d(e^{-x}) \\
&= -e^{-x}x^n|_{\mu}^{\infty} + \int_{x=\mu}^{\infty} e^{-x} d(x^n) \quad [\text{ by parts}] \\
&= e^{-\mu}\mu^n + n\int_{\mu}^{\infty} e^{-x}x^{n-1}dx \quad [\because e^{-x}x^n \to 0 \text{ as } x \to \infty] \\
&= e^{-\mu}\mu^n + nI_{n-1}.
\end{aligned}$$

এই জিনিসটা যতই লাগাব ততই $I_{n-1}, I_{n-2}, \ldots$ এইভাবে চলবে। আমাদের অংকে অবশ্য I_n-কে $n!$ দিয়ে ভাগ করা ছিল। এবার সেইটাকে হিসেবে আনি।

So

$$J_n := \frac{I_n}{n!} = \frac{e^{-\mu}\mu^n + nI_{n-1}}{n!} = \frac{e^{-\mu}\mu^n}{n!} + \frac{I_{n-1}}{(n-1)!} = \frac{e^{-\mu}\mu^n}{n!} + J_{n-1} = P(X=n) + J_{n-1}.$$

Using this repeatedly

$$J_n = P(X=n) + P(X=n-1) + J_{n-2} = \cdots = P(X=n) + \cdots + P(X=1) + J_0.$$

Now

$$J_0 = \int_\mu^\infty e^{-x}dx = e^{-\mu} = P(X=0).$$

So $J_n = P(X \leq n)$, as required.

■

Example 21: If m and μ_k denote the mean and k-th central moment of a Poisson distribution, then prove that

$$\mu_{k+1} = m\left(k\mu_{k-1} + \frac{d\mu_k}{dm}\right).$$

Hence find the coefficient of skewness γ_1 and the coefficient of excess of kurtosis γ_2 of this distribution.[5] **(2012.3c)**

SOLUTION: এই অংকে বলতে ভুলে গেছে যে m-টা কী! আসলে এখানে আমরা Poisson(m) নিয়ে কাজ করছি।

If $X \sim \text{Poisson}(m)$, then $E(X) = m$.

$\therefore$ By definition

$$\mu_k = E\big((X-m)^k\big) = \sum_0^\infty e^{-m}(x-m)^k \frac{m^x}{x!}.$$

এবার এটাকে m-এর সাপেক্ষে differentiate করব।

So

$$\begin{aligned}\frac{d\mu_k}{dm} &= \sum_{x=0}^{\infty} \underbrace{\big(-e^{-m}\big)}(x-m)^k \frac{m^x}{x!} \\ &\quad \sum_{x=0}^{\infty} e^{-m} \underbrace{(-k)(x-m)^{k-1}} \frac{m^x}{x!} \\ &\quad + \sum_{x=0}^{\infty} e^{-m}(x-m)^k \underbrace{\frac{xm^{x-1}}{x!}} \\ &= -\mu_k - k\mu_{k-1} + \sum_{x=0}^{\infty} e^{-m}(x-m)^k \frac{xm^{x-1}}{x!}.\end{aligned}$$

এবার শেষের sum-টাকে দলাইমলাই করব।

Now

$$\begin{aligned}&\sum_{x=0}^{\infty} e^{-m}(x-m)^k \frac{xm^{x-1}}{x!} \\ =& \sum_{x=0}^{\infty} e^{-m}(x-m)^k \underbrace{(x-m+m)}_{x} \frac{m^{x-1}}{x!}\end{aligned}$$

লক্ষ কর x-টাকে কীভাবে $x-m+m$ আকারে লিখেছি।

এবার $x-m+m$-কে দুভাগে ভাগ করব, $x-m$ আর m. ফলে দুটো sum পাব--

$$\begin{aligned}=& \sum_{x=0}^{\infty} e^{-m}(x-m)^{k+1} \frac{m^{x-1}}{x!} + \sum_{x=0}^{\infty} e^{-m}(x-m)^k \frac{m^x}{x!} \\ =& \frac{1}{m}\mu_{k+1} + \mu_k.\end{aligned}$$

So

$$\begin{aligned}\frac{d\mu_k}{dm} &= -\mu_k - k\mu_{k-1} + \frac{1}{m}\mu_{k+1} + \mu_k \\ &= -k\mu_{k-1} + \frac{1}{m}\mu_{k+1}.\end{aligned}$$

So

$$\mu_{k+1} = m\left(k\mu_{k-1} + \frac{d\mu_k}{dm}\right), \qquad (*)$$

as required.

এবার দ্বিতীয় অংশ। এখানে coefficient of skewness আর coefficient of excess kurtosis বার করতে হবে, মানে $\mu_3/(\mu_2)^{3/2}$ আর $\mu_4/(\mu_2)^2$. সুতরাং μ_2, μ_3 আর μ_4 বার করতে হবে।

We know that $\mu_1 = 0$ and $\mu_2 = m$ for a Poisson(m) distribution. So putting $k = 2$ in (*)

$$\mu_3 = m\left(2\mu_1 + \frac{d\mu_2}{dm}\right) = m(2 \times 0 + 1) = m.$$

Also putting $k = 3$,

$$\mu_4 = m\left(3\mu_2 + \frac{d\mu_3}{dm}\right) = m(2m + 1).$$

So coefficient of skewness is

$$\frac{\mu_3}{(\mu_2)^{3/2}} = \frac{m}{m^{3/2}} = \frac{1}{\sqrt{m}},$$

and coefficient of excess of kurtosis is

$$\frac{\mu_4}{(\mu_2)^2} - 3 = \frac{m(2m+1)}{m^2} - 3 = \frac{1}{m} - 1.$$

■

Example 22: Let λ and μ_r denote the mean and r-th moment about mean of a Poisson distribution respectively. Obtain the following recurrence relation

$$\mu_{r+1} = r\lambda\mu_{r-1} + \lambda\frac{d\mu_r}{d\lambda}.$$

(2005.3c)
SOLUTION: আগের অংকের প্রথম অংশ। ■

17.2 Geometric distribution

ধরো একটা Bernoulli random experiment আছে। যথারীতি এর দুটো outcome, যাদের নাম দিলাম success আর failure. যদি এটাকে n বার করি তবে success-এর মোট সংখ্যা যে একটা binomial random variable হবে সেটা আমরা শিখেছি। বাস্তবে একইরকম ব্যাপার অনেক সময়ে একটু অন্যরূপে দেখা দেয়, যার ফলে একটা অন্য probability distribution-এর জন্ম হয়। একটা উদাহরণ দেখি।

Example 23: তুমি মোবাইলে ফোন করতে চাও। এর জন্য টাওয়ার পাওয়া দরকার। এই "টাওয়ার পাওয়া" ব্যাপারটা

এইরকম-- তোমার মোবাইলটা কিছুক্ষণ বাদেবাদে একটা সিগনাল পাঠাতে থাকে। যদি কাছাকাছির মধ্যে কোনো টাওয়ার থাকে তবে সেটা তোমার সিগনালকে চিনতে পারে, এবং উত্তর হিসেবে আরেকটা সিগনাল পাঠায়, যেটা আবার তোমার মোবাইল চিনতে পারে। এই দ্বিপাক্ষিক চেনাচিনির পর্ব সাঙ্গ হলেই আমরা উল্লসিত হয়ে বলি "টাওয়ার পেয়েছে!" মোবাইল থেকে টাওয়ারে এবং টাওয়ার থেকে মোবাইলে সিগনাল পৌঁছনোর পথে নানা সমস্যা হতে পারে--দূরত্ব বেশী, বা সিগনাল দুর্বল, বা আবহাওয়ার গোলযোগ ইত্যাদি। সুতরাং টাওয়ার পাওয়া না পাওয়াটা খানিকটা ভাগ্যের উপর নির্ভর করে। বুঝতেই পারছ এটা একটা random experiment. যেহেতু এর দুটো মাত্র outcome সম্ভব--"পাওয়া" (success) এবং "না পাওয়া"

(failure) তাই এটা একটা Bernoulli random experiment-ও বটে। তুমি যখন ফোন করতে চাও, মোবাইলটা কিন্তু এই random experiment-টা একবার করেই হাল ছেড়ে দেয় না। বার বার করে চেষ্টা করতেই থাকে, যতক্ষণ না টাওয়ার পায়। একবার পেয়ে গেলে আর চেষ্টা করে না বলাই বাহুল্য। এই অবস্থায় মোট কতবার ব্যর্থ হবার পর প্রথম সাফল্য পাওয়া গেল সেটা একটা random variable. এটা একটা গুরুত্বপূর্ণ জিনিস, কারণ এর উপর অনেক কিছু নির্ভর করে--কত সময় লাগছে, কত ব্যাটারী পুড়ছে ইত্যাদি। সুতরাং এর probability distribution-টা বার করা দরকারী। এই probability distribution-টার নাম হল **geometric distribution.** এর p.m.f. বার করতে বলেছে নীচের অংকে। ■

Example 24: Find the probability distribution of number of failures preceding the first success in an infinite sequence of Bernoulli's trial with probability of success p for each trial.[5] **(2014,2004)**

SOLUTION: এই অংকের সাথে মোবাইলের টাওয়ার পাওয়ার গল্পটা মিলিয়ে নাও। এখানে "infinite sequence" বলার অর্থ মোবাইলটা কোনোদিনই হাল ছাড়বে না, চিরকাল চেষ্টা চালিয়েই যাবে[2]। প্রত্যেক চেষ্টার পর টাওয়ার পাওয়ার probability বলে দিয়েছে p.

Let X be the number of failures proceeding the first success.

Then clearly, X can take values $0, 1, 2, ...$

$X = 0$ মানে প্রথম চেষ্টাতেই বাজিমাত, একবারও ব্যর্থ হতে হয় নি। $X = 1$ মানে প্রথমবার ব্যর্থতা, কিন্তু দ্বিতীয় চেষ্টাতেই সাফল্য।

Let $x \in \{0, 1, 2, ...\}$.

Then

$$\begin{aligned} P(X = x) &= P(x \text{ failures followed by a success}) \\ &= (1-p)^x p \quad [\because \text{independent}] \end{aligned}$$

Thus the required probability distribution has p.m.f.

$$P(X = x) = \begin{cases} (1-p)^x p & \text{if } x = 0, 1, 2, ... \\ 0 & \text{otherwise} \end{cases}$$

■

একে কেন geometric বলে বুঝতে পারছ নিশ্চয়ই, $P(X = 0), P(X = 1), P(X = 2), ...$ ইত্যাদিরা একটা GP-তে রয়েছে। এবার আমরা সংজ্ঞাটা গুছিয়ে লিখে ফেলি--

DEFINITION: **Geometric distribution**

Let X be a random variables taking only nonnegative integer values. It is said to have a Geometric(p) distribution, if the PMF is

$$P(X = x) = \begin{cases} (1-p)^x p & \text{if } x = 0, 1, 2, ... \\ 0 & \text{otherwise} \end{cases}$$

Here $p \in (0, 1)$ is a parameter. We write $X \sim \text{Geometric}(p)$.

[2] খুব জোরালো ব্যাটারী বলতে হবে!

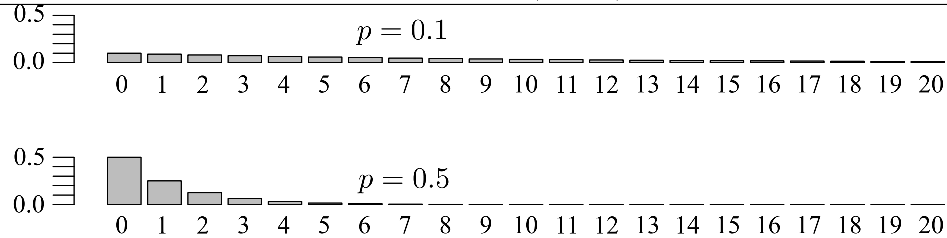


Fig 5

এই family-র কয়েকটা PMF দেখিয়েছি Fig 5-এ।

DAY 18
Standard continuous distribution families (part 1)

গত দুদিন আমরা বিভিন্ন discrete distribution নিয়ে আলোচনা করেছি। আজকে আর কালকে আমরা কিছু continuous distribution নিয়ে আলোচনা করব। এদের প্রত্যেকেরই PDF থাকবে।

18.1 Normal distribution

এটা হল statistics-এর দুনিয়ায় সবচেয়ে বহুল ব্যবহৃত distribution. এর সঙ্গে আমাদের আগেই পরিচয় হয়েছে। সংজ্ঞাটা আরেকবার মনে করিয়ে দিই।

DEFINITION: **Normal/Gaussian distribution**

A random variable X is said to have **normal/Gaussian distribution** if it has PDF

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad x \in \mathbb{R}.$$

Here $\mu \in \mathbb{R}$ and $\sigma > 0$ are two parameters. We write $X \sim N(\mu, \sigma)$. The special case where $\mu = 0$ and $\sigma = 1$ is called the **standard normal distribution**.

বিভিন্ন normal distribution-এর PDF-এর ছবি দেখতে পাবে Fig 6-এ। ছবিগুলোকে অনেকটা অ্যানিমেশনের মত ভাবতে পারো। যদি μ বদলাও, তবে গ্রাফটার চেহারা বদলাবে না, খালি পাশাপাশি সরবে। যদি σ বাড়াও তবে গ্রাফটা একই জায়গায় থাকবে, কিন্তু বেশী ছড়িয়ে যাবে। যদি σ কমাও, তবে একই জায়গায় থেকে আরো সরু আর উঁচু হয়ে যাবে।
এবার বলি এই distribution-এর mean এবং variance কত।

THEOREM

If $X \sim N(\mu, \sigma)$, then

$$E(X) = \mu \text{ and } \operatorname{var}(X) = \sigma^2.$$

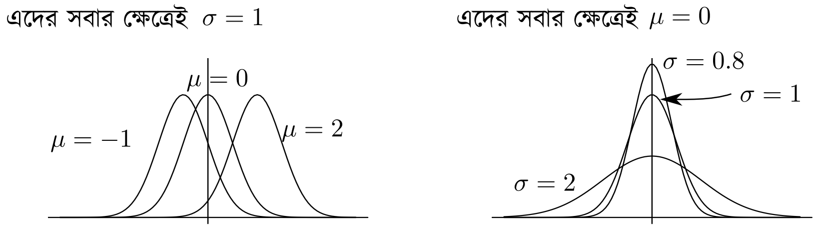


Fig 6

নীচের অংকটায় এগুলোই বার করে দেখাতে বলেছে।

Example 25: Calculate mean and variance of Gaussian distribution.[2+3] **(2010,2009)**

SOLUTION:

Let $X \sim N(\mu, \sigma)$ for some $\mu \in \mathbb{R}$ and $\sigma > 0$.
Then the p.d.f. of X is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad x \in \mathbb{R}.$$

So

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma t + \mu) \exp\left[-\frac{t^2}{2}\right] dt \quad \left[\text{putting } t = \tfrac{x-\mu}{\sigma}\right] \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t \exp\left[-\frac{t^2}{2}\right] dt + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{t^2}{2}\right] dt \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t \exp\left[-\frac{t^2}{2}\right] dt + \mu, \end{aligned}$$

Since $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = 1$, because this is the integral of $N(0,1)$ PDF.

এবার অন্য integral-টার দিকে তাকানো যাক--

Now $t \exp\left[-\frac{t^2}{2}\right]$ is an odd function.

So the first integral is 0 if $\int_0^{\infty} t \exp\left[-\frac{t^2}{2}\right] dt < \infty$.

এই জায়গাটা একটু বুঝে নেওয়া যাক। এখানে যে odd function-টাকে integrate করছি সেটা হল $te^{-t^2/2}$. গ্রাফে দেখতে Fig 7-এর মত। এখানে $t > 0$ আর $t < 0$-র জন্য যে দুটো area পাচ্ছি তারা ঠিক একইরকমের, খালি একজন উল্টে

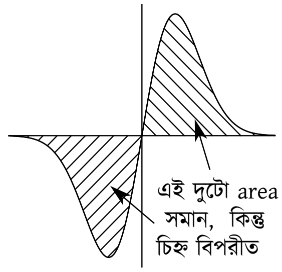


Fig 7

আছে, এই যা পার্থক্য। তাই মোট integral-টা হবে উপরের area থেকে নীচের area-টা বিয়োগ করলে যা হয় সেটা। যেহেতু দুটো area-ই সমান তাই উত্তরটা 0 হওয়াই উচিত। খালি একটা সমস্যা আছে, যদি area-টা ∞ হয়ে যায়, তবে কিন্তু বিয়োগ করলে $\infty - \infty$ জাতীয় একটা জিনিস চলে আসবে, যেটা undefined. সেই সমস্যা যে এখানে আসলে হবে না, সেটা দেখানোর জন্যই $\int_0^\infty te^{-t^2/2}dt < \infty$ প্রমাণ করা দরকার।

But

$$\begin{aligned}\int_0^\infty t\exp\left[-\frac{t^2}{2}\right]dt &= \int_{t=0}^\infty \exp\left[-\frac{t^2}{2}\right]d\left(\frac{t^2}{2}\right)\\ &= \int_0^\infty e^{-u}du \quad \left[\text{putting } u = t^2/2\right]\\ &= 1.\end{aligned}$$

So $E(X) = \mu$.

এবার $\text{var}(X)$ বার করা পালা--

Now

$$\begin{aligned}\text{var}(X) &= E(X - E(X))^2 = E(X-\mu)^2\\ &= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^\infty (x-\mu)^2\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]dx\\ &= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^\infty t^2\exp\left[-\frac{t^2}{2}\right]dt \quad \left[\text{putting } t = \tfrac{x-\mu}{\sigma}\right]\\ &= \frac{2\sigma^2}{\sqrt{2\pi}}\int_0^\infty t^2\exp\left[-\frac{t^2}{2}\right]dt \quad [\because \text{even function}]\\ &= \frac{2\sigma^2}{\sqrt{\pi}}\int_0^\infty u^{1/2}e^{-u}du \quad \left[\text{putting } u = \tfrac{t^2}{2}\right]\end{aligned}$$

এই ধাপটা ভালো করে নিজে লিখে বুঝে নাও।

$$= \frac{2\sigma^2}{\sqrt{\pi}}\Gamma\left(\frac{3}{2}\right)$$

মনে রেখো যে $\Gamma(a) = \int_0^\infty u^{a-1}e^{-u}du.$

$$= \frac{2\sigma^2}{\sqrt{\pi}}\frac{1}{2}\Gamma\left(\frac{1}{2}\right) \quad [\because \forall b > 0 \quad \Gamma(b+1) = b\Gamma(b)]$$

$$= \frac{\sigma^2}{\sqrt{\pi}}\sqrt{\pi} = \sigma^2.$$

■

Example 26: The probability density function of a random variable is given by

$$f(x) = ce^{-(x^2+2x+3)}, \quad -\infty < x < \infty.$$

Find the value of the constant c, the expectation and variance of the distribution. **(2013.3b)**

SOLUTION:

We know that any density integrates to 1.

$$\therefore \int_{-\infty}^{\infty} f(x)dx = c\int_{-\infty}^{\infty} e^{-(x^2+2x+3)}dx = 1,$$

ie,

$$c = \frac{1}{\int_{-\infty}^{\infty} e^{-(x^2+2x+3)}dx}. \qquad (*)$$

Now

$$\int_{-\infty}^{\infty} e^{-(x^2+2x+3)}dx = \int_{-\infty}^{\infty} \exp(-(x+1)^2 - 2)dx$$

$$= e^{-2}\int_{-\infty}^{\infty} \exp\left(-\frac{(x+1)^2}{2\times(1/\sqrt{2})^2}\right)dx.$$

এরকম অদ্ভুতভাবে লেখার কারণ আর কিছুই নয়, পুরোটাকে normal distribution-এর PDF-এর মত করে লিখছি। তাহলে integration-টা আর কষ্ট করে করতে হবে না। কী করে দ্যাখো--

We know that $N(\mu,\sigma)$ density is

$$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

So

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx = 1,$$

> or,
> $$\int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \sqrt{2\pi}\sigma.$$

আমাদের যে integration-টা করতে হবে, সেটাও ঠিক এই চেহারার। সুতরাং--

> Taking $\mu = -1$ and $\sigma = 1/\sqrt{2}$, we have
> $$e^{-2}\int_{-\infty}^{\infty} \exp\left(-\frac{(x+1)^2}{2\times(1/\sqrt{2})^2}\right) dx = \sqrt{\pi}e^{-2}.$$
> Thus, by (*),
> $$c = \frac{e^2}{\sqrt{\pi}}.$$
> Also with this choice, $f(x)$ is just $N\left(-1, \frac{1}{\sqrt{2}}\right)$ density.
>
> So the expectation is $\mu = -1$ and variance is $\sigma^2 = \frac{1}{2}$.

■

Example 27: Show that the mean deviation about the mean of a normal (m, σ) distribution is $\sqrt{\frac{2}{\pi}}\sigma$.[5] **(2011.2e)**

SOLUTION: এখানে একটা নতুন কথা আছে, mean deviation about mean. এর মানে হল $E\big|X - E(X)\big|$, মানে আমাদের অংকে $E|X - m|$.

> $$\begin{aligned} E|X-m| &= \int_{-\infty}^{\infty} |x-m| \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma|t| \exp\left(-\frac{t^2}{2}\right) dt \quad \left[\text{putting } t = \tfrac{x-m}{\sigma}\right] \\ &= \frac{2\sigma}{\sqrt{2\pi}} \int_0^{\infty} t \exp\left(-\frac{t^2}{2}\right) dt \quad [\because \text{even function}] \\ &= \frac{2\sigma}{\sqrt{2\pi}} \int_0^{\infty} e^{-u} du \quad \left[\text{putting } u = t^2/2\right] \\ &= \frac{2\sigma}{\sqrt{2\pi}} = \sqrt{\frac{2}{\pi}}\sigma, \end{aligned}$$
>
> as required.

■

Exercise 4: Show that the first absolute moment about the mean for the normal(m, σ) distribution is

$$\sqrt{\frac{2}{\pi}}\sigma.$$

(2007.2d)

HINT:

আগের অংকটাই, কারণ first absolute moment about the mean মানেও সেই $E\big|X - E(X)\big|$. ■

এবার শিখব $N(\mu, \sigma)$-এর MGF আর characteristic function কী।

> THEOREM
>
> The MGF of $N(\mu, \sigma)$ distribution is
>
> $$\phi(t) = \exp\left(\mu t + \tfrac{1}{2}\sigma^2 t^2\right), \quad t \in \mathbb{R}.$$
>
> $$Q(t) = \exp\left(i\mu t - \tfrac{1}{2}\sigma^2 t^2\right), \quad t \in \mathbb{R}.$$

MGF-এর প্রমাণটা চেয়েছে নীচের অংকে। একইভাবে characteristic function-এর প্রমাণটাও হবে।

Example 28: Calculate the m.g.f. for a normal(m, σ) variate.[6] **(2005.4a)**

SOLUTION:

Let $X \sim N(m, \sigma)$. Then the required m.g.f. is

$$
\begin{aligned}
E(e^{tX}) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp(tx)\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)dx \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\left(-\frac{(x-m)^2}{2\sigma^2}+tx\right)dx \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\left(-\frac{(x-m)^2-2\sigma^2 tx}{2\sigma^2}\right)dx \\
&= \cdots
\end{aligned}
$$

এখানে তোমার কাজ হবে square complete করা,
মানে ওই যে $(x-m)^2 - 2\sigma^2 tx$ আছে, ওটাকে তুমি
"$(x-$ কিছু একটা$)^2+$ কিছু একটা" আকারে প্রকাশ করবে।
"কিছু একটা"-গুলোর মধ্যে যেন কোনো x না থাকে। তাহলে পাবে--

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\left(-\frac{(x-m-\sigma^2 t)^2 + m^2 - (m+\sigma^2 t)^2}{2\sigma^2}\right)dx \\
&= \exp\left(\frac{m^2-(m+\sigma^2 t)^2}{-2\sigma^2}\right)\underbrace{\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\exp\left(-\frac{(x-m-\sigma^2 t)^2}{2\sigma^2}\right)dx}_{1}
\end{aligned}
$$

এবার লক্ষ কর যে, তলায় দাগ দিয়ে যেটাকে 1 লিখেছি, সেটা কেন 1.
আসলে ওটা হল $N(m+\sigma^2 t, \sigma)$-র PDF-এর integral.

$$
\begin{aligned}
&= \exp\left(\frac{m^2-(m+\sigma^2 t)^2}{-2\sigma^2}\right) \\
&\qquad \left[\because \text{total integral of } N(m+\sigma^2 t, \sigma) \text{ PDF} = 1\right] \\
&= \exp\left(\frac{2m\sigma^2 t + \sigma^4 t^2}{2\sigma^2}\right) \\
&= \exp\left(mt + \tfrac{1}{2}\sigma^2 t^2\right) \text{ for } t \in \mathbb{R}.
\end{aligned}
$$

■

এবার একটা খুব গুরুত্বপূর্ণ জিনিস শিখব, যা দিয়ে যেকোনো $N(\mu, \sigma)$ থেকেই standard normal distribution (মানে $N(0,1)$-এ) পৌঁছনো যায়।

THEOREM

If $X \sim N(\mu, \sigma)$ for any $\mu \in \mathbb{R}$ and any $\sigma > 0$ then

$$\frac{X-\mu}{\sigma} \sim N(0,1).$$

এর প্রমাণ খুবই সহজ। কিন্তু তার আগে একটা নাম শিখে রাখি। এখানে আমরা X-এর থেকে μ বিয়োগ করলাম, মানে $E(X)$ বিয়োগ করলাম, তারপর σ দিয়ে ভাগ করলাম, মানে $\sqrt{\text{var}(X)}$ দিয়ে ভাগ করলাম। এই কাজটাকে বলে **standardisation.** শুধু normal-এর ক্ষেত্রেই নয়, যেকোনো random variable-এর ক্ষেত্রেই এই ভাষা চলে। যদি normal হয়, তবে standardisation-এর ফলে সবসময়ে $N(0,1)$ distribution পাওয়া যায়। এটাই theorem-টার বক্তব্য।

প্রমাণটা অনায়াসে হবে characteristic function দিয়ে। ধরো $X \sim N(\mu,\sigma)$. তাহলে তার characteristic function হবে

$$Q(t) \equiv E\big(e^{itX}\big) = \exp\left(i\mu t - \tfrac{\sigma^2 t^2}{2}\right).$$

এবার ধরো $Y = \frac{X-\mu}{\sigma}$ নিলাম। তার characteristic function হবে

$$E\big(e^{itY}\big) = E\left(e^{\frac{it(X-\mu)}{\sigma}}\right) = e^{\frac{-i\mu t}{\sigma}} E\left(e^{\frac{itX}{\sigma}}\right) = e^{\frac{-i\mu t}{\sigma}} Q\left(\tfrac{t}{\sigma}\right) = \cdots = \exp\left(-\tfrac{t^2}{2}\right).$$

মাঝখানে ডট্‌ডটের জায়গায় কটা মামুলী ধাপ আছে, ওটা তোমার জন্যে রেখে দিলাম।

Exercise 5: If X is a standard normal variate, find the mean and variance of e^X.[5] **(2012,2009)**

HINT:

এই অংকটা X-এর MGF ব্যবহার করে বেশ সহজে হবে। ধরো $Y = e^X$. আমাদের বার করতে বলেছে $E(Y)$ আর $\text{var}(Y) = E(Y^2) - \big(E(Y)\big)^2$. এখানে $E(Y) = E(e^X) = \phi(1)$, যেখানে $\phi(t)$ হল X-এর MGF. একইভাবে $E(Y^2) = E(e^{2X}) = \phi(2)$. যেহেতু standard normal distribution-এর MGF জানা আছে, তাই $\phi(1)$ আর $\phi(2)$ বার করতে আর কী আছে? সুতরাং অনায়াসে উত্তর বেরিয়ে যাবে। চেষ্টা করেই দ্যাখো! ■

Example 29: If the independent random variables $X_1, ..., X_n$ all have the same distribution and their sum is normally distributed, then prove that each of them is normally distributed.[6] **(2004)**

SOLUTION: এই একটা অংক যেটা characteristic function ছাড়া করা শক্ত।

> Let the characteristic function of each X_j be $\phi(t)$.

লক্ষ কর যে, ϕ-এর মধ্যে কোনো j নেই, কারণ সবগুলো X_j-এরই একই distribution, তাই characteristic function-ও একই।

> Since the X_j's are independent, their sum has characteristic function
>
> $$\underbrace{\phi(t) \times \cdots \times \phi(t)}_{n \text{ times}} = (\phi(t))^n.$$
>
> Let the sum have $N(\mu,\sigma)$ distribution.
> Then, matching characteristic functions,
>
> $$(\phi(t))^n = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right).$$
>
> So $\phi(t) = \exp\left(\frac{i\mu t}{n} - \frac{\sigma^2 t^2}{2n}\right)$, which is the characteristic function of $N\left(\frac{\mu}{n}, \frac{\sigma}{\sqrt{n}}\right)$

> distribution.
> Hence, by characterising property of characteristic functions, each X_i has this normal distribution, as required.

■

এইবার একটা অংক দেব, যেটা খানিকটা কসরতের, কিন্তু পরে কোথাও কাজে লাগবে না। পরীক্ষার চিন্তা না থাকলে চাইলে বাদ দিয়ে যেতে পার।

Exercise 6: If X is a normal(m, s) variate, and μ_k denotes the k-th central moment of X, then prove that

$$\mu_{2k+2} = \sigma^2 \mu_{2k} + \sigma^3 \frac{d\mu_{2k}}{d\sigma}.$$

[4] **(2004.4c)**
HINT:
এই অংকটা পুরো করে দেব না, খালি মূল কায়দাটা ধরিয়ে দিচ্ছি। এখানে

$$\mu_{2k} = \int_{-\infty}^{\infty} \frac{(x-m)^{2k}}{\sqrt{2\pi}\sigma} \exp\left(\tfrac{(x-m)^2}{2\sigma^2}\right) dx.$$

যদি দাঁতে দাঁত চেপে integrate করে যেতে পারো, তবে যেটা প্রমাণ করতে বলেছে তার RHS থেকে শুরু করে এক সময়ে LHS-এ পৌঁছবে। তবে কিনা দাঁতটা এত জোরে চাপতে হবে যে, দাঁত ভেঙে যেতে পারে। একটু সহজ কায়দা হল প্রথমে standardise করে নেওয়া--

> Let $Z = \frac{X-m}{\sigma} \sim N(0,1)$.
> For $\ell = 1, 2, 3, ...,$ let $r_\ell = E(Z^\ell)$.
> Then $\mu_\ell = E\big((X-m)^\ell\big) = \sigma^\ell E(Z^\ell) = \sigma^\ell r_\ell$.
> We want to show
> $$\mu_{2k+2} = \sigma^2 \mu_{2k} + \sigma^3 \frac{d\mu_{2k}}{d\sigma},$$
> ie,
> $$\sigma^{2k+2} r_{2k+2} = \sigma^2 \sigma^{2k} r_{2k} + \sigma^3 \frac{d(\sigma^{2k} r_{2k})}{d\sigma},$$

মনে রেখো যে, r-গুলোর মধ্যে কোনো σ-টিগ্‌মা নেই, ওরা স্রেফ constant. সুতরাং শেষের ওই differentiation-টা খুবই সহজ। সব কেটেকুটে পাবে

> ie,
> $$r_{2k+2} = r_{2k} + 2k\, r_{2k}. \qquad (*)$$

এইটা অনেকটাই বেশী ভদ্র দেখতে। বিশেষতঃ, এখন আর কোনো m বা σ নিয়ে মাথা ঘামাতে হবে না। এইবার খালি r_{2k}-র একটা ফর্মুলা বার করতে হবে integrate করে। ওই ফর্মুলাতে k-র জায়গায় $k+1$ বসিয়ে দিলেই r_{2k+2}-ও বেরিয়ে যাবে। অমনি দেখবে (*)-এর দুই পাশ দিব্যি মিলে যাচ্ছে। সুতরাং কঠিন কাজ খালি একটাই, r_{2k}-র ফর্মুলা বার করা। ফর্মুলাটা কী হবে বলে দিচ্ছি, কষে মিলিয়ে নিও।

Here $\mu_{2k} = \cdots = \frac{1}{\sqrt{\pi}}[2^k\Gamma\left(k+\frac{1}{2}\right)]$.

বাকী লেখাটুকু তোমার দায়িত্ব। ■

DAY 19
Standard continuous distribution families (part 2)

গতকাল আমরা normal distribution নিয়ে অনেক কথা শিখেছি। আজকে আরও কয়েকটা continuous distribution-এর কথা শিখব।

19.1 Exponential distribution

এই distribution-টার সঙ্গে আমাদের আগেই পরিচয় হয়েছে। সংজ্ঞাটা ছিল এইরকম--

DEFINITION: Exponential distribution

A random variable X is said to have **exponential distribution with rate** λ if it has PDF

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\lambda > 0$ is a parameter. We write $X \sim \text{Exponential}(\lambda)$.

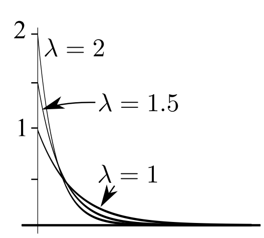
কিছু exponential distribution-এর PDF রয়েছে Fig 8-এ।
অতি সহজ integration থেকেই পেয়ে যাবে এর mean আর variance–

THEOREM

If $X \sim \text{Exponential}(\lambda)$ for some $\lambda > 0$, then

$$E(X) = \frac{1}{\lambda} \text{ and } \text{var}(X) = \frac{1}{\lambda^2}.$$

Fig 8



Example 30: Find the median and mode, if any, of the distribution having probability density function

$$f(x) = \lambda e^{-\lambda x} \quad x > 0.$$

[5] **(2008.2b)**

SOLUTION: প্রথমে median-এর সংজ্ঞাটা মনে করে নিই--

> We know that for a continuous random variable X the median is defined as any number c such that $P(X \leq c) = \frac{1}{2}$.
>
> Let X have the given PDF.
>
> Let $c \in \mathbb{R}$.
>
> If $c < 0$, then $P(X \leq c) = 0$.
>
> If $c \geq 0$, then $P(X \leq c) = \int_0^c \lambda e^{-\lambda x} dx = 1 - e^{-\lambda c}$.
>
> Thus, $P(X \leq c) = \frac{1}{2} \iff 1 - e^{-\lambda c} = \frac{1}{2}$, or $c = \frac{\log 2}{\lambda}$.
>
> Hence the median of the given distribution is $c = \frac{\log 2}{\lambda}$.

■

19.2 Gamma distribution

DEFINITION: Gamma distribution

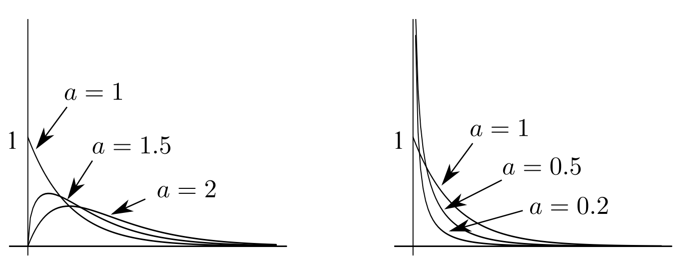
A random variable X is said to have **gamma distribution** if it has probability density function

$$f(x) = \begin{cases} \frac{1}{\Gamma(a)} x^{a-1} e^{-x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Here $a > 0$ is a parameter. We write $X \sim \text{Gamma}(a)$.

এই family-র কিছু সদস্যের PDF-এর গ্রাফ এঁকেছি Fig 9-এ।

Fig 9



THEOREM

If $X \sim \text{Gamma}(a)$ for some $a > 0$, then

$$E(X) = a \text{ and } \text{var}(X) = a.$$

এগুলো সরাসরি integrate করেই দেখানো যায়। নীচের অংকে অবশ্য characteristic function দিয়ে দেখাতে বলেছে।

Example 31: Find the characteristic function of Gamma(1) distribution and hence find its mean and variance.[3+1+1] **(2012.3b)**

SOLUTION:

The density of Gamma(1) distribution is

$$f(x) = \begin{cases} e^{-x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

So the required characteristic function is

$$\begin{aligned} Q(t) &= E(e^{itX}) \\ &= \int_{-\infty}^{\infty} e^{itx} f(x) dx \\ &= \int_{0}^{\infty} e^{itx} e^{-x} dx \\ &= \int_{0}^{\infty} e^{(it-1)x} dx \\ &= \frac{1}{it-1} e^{(it-1)x}\Big|_0^\infty \\ &= \frac{1}{1-it}, \quad t \in \mathbb{R}, \end{aligned}$$

Since e^{it} is bounded (for $t \in \mathbb{R}$), and $E^{-x} \to 0$ as $x \to \infty$.

এবার চট করে মনে করে নিই, characteristic function ব্যবহার করে কীকরে বিভিন্ন raw moment বার করে। যদি X-এর characteristic function হয় $Q(t)$, তবে $E(X^k)$ হবে $Q^{(k)}(0)/i^k$, অর্থাৎ $Q(t)$-কে প্রথমে k-বার differentiate করে, তাতে $t = 0$ বসাতে হবে। যে উত্তরটা পাওয়া যাবে তাকে i^k দিয়ে ভাগ করতে হবে।
আমাদের বার করতে বলেছে mean আর variance. তাই আমাদের লাগবে $E(X)$ আর $E(X^2)$.

So

$$E(X) = \frac{\phi'(0)}{i} = \frac{1}{i} \times \frac{i}{(1-it)^2}\Big|_{t=0} = 1.$$

Similarly,

$$E(X^2) = \frac{\phi''(0)}{i^2} = \frac{1}{i^2} \times \frac{-2}{(1-it)^3}\Big|_{t=0} = 2.$$

> So $var(X) = E(X^2) - (E(X))^2 = 2 - 1 = 1.$

■

এই অংকে আমরা Gamma(1)-এর characteristic function বার করলাম। একইভাবে Gamma(n)-এর characteristic function-ও বার করা যেত। সেখান থেকেই এই theorem-টার জন্ম--

THEOREM

The characteristic function of Gamma(n) distribution is

$$Q(t) = (1 - it)^{-n}, \quad t \in \mathbb{R}.$$

Example 32: If a random variable X is a gamma(1) variate and Y is a gamma(m) variate, show that $X + Y$ is a gamma($1 + m$) variate, where X, Y are independent.[3] **(2014.3c)**

SOLUTION:

> We know that the characteristic function of Gamma(n) distribution is $Q(t) = (1 - it)^{-n}$.
>
> So the characteristic function of X and Y are, respectively,
>
> $$Q_X(t) = (1 - it)^{-1},$$
> $$Q_Y(t) = (1 - it)^{-m}.$$
>
> Since X, Y are independent, the characteristic function of $X + Y$ is
>
> $$Q_X(t)Q_Y(t) = (1 - it)^{-1}(1 - it)^{-m} = (1 - it)^{-(m+1)},$$
>
> which is the characteristic function of Gamma($m + 1$) distribution.
>
> Since characteristic function characterises a distribution, hence $X + Y$ must have Gamma($m + 1$) distribution, as required.

■

19.3 Beta distribution

এবার আমরা beta distribution নামে একধরণের distribution-এর কথা শিখব। এদের খালি সংজ্ঞা দিয়েই ছেড়ে দেব, কোনো বিশদ আলোচনায় যাব না।

প্রথমে Beta function-এর কথা একটু মনে করিয়ে দিই। আমাদের বাংলায়-বোঝানো-ইংরাজি-বই সিরিজের Real Analysis (volume 3)-তে এর বিষয়ে আলোচনা আছে। সংজ্ঞাটা ছিল এইরকম--

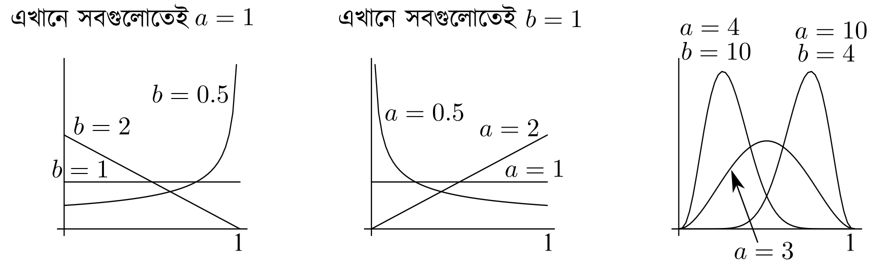


Fig 10

> DEFINITION: **Beta function**
>
> The **Beta function** is defined as
>
> $$\text{Beta}(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt, \quad x, y > 0.$$

এই function-টাকে ব্যবহার করেই beta distribution-দের জন্ম।

> DEFINITION: **Beta distribution (type 1)**
>
> A random variable X is said to have **beta distribution (type 1)** if it has probability density function
>
> $$f(x) = \begin{cases} \frac{1}{\text{Beta}(a,b)} x^{a-1}(1-x)^{b-1} & \text{if } x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$
>
> where $a, b > 0$ are two parameters. We write $X \sim \beta_1(a, b)$.

> DEFINITION: **Beta distribution (type 2)**
>
> A random variable X is said to have **beta distribution (type 2)** if it has probability density function
>
> $$f(x) = \begin{cases} \frac{1}{\text{Beta}(a,b)} \frac{x^{a-1}}{(1+x)^{a+b}} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$
>
> where $a, b > 0$ are two parameters. We write $X \sim \beta_2(a, b)$.

দুই ধরণের beta distribution থাকলেও, এদের মধ্যে প্রথমটাই বেশী ব্যবহার হয়। সুতরাং যদি কোথাও দ্যাখো খালি beta distribution লিখেছে (কোনো type-এর উল্লেখ ছাড়া) তবে জানবে যে, type 1-এর কথাই বলা হচ্ছে। Beta (type 1) distribution -এর কিছু PDF দেখতে পাচ্ছ Fig 10-এ।

Beta distribution-দেরও mean, variance, characteristic function ইত্যাদি আছে। কিন্তু আমরা সে সবের আলোচনায় যাব না।

Answers

2. $\left(pe^t + (1-p)\right)^n$. **5.** $E(e^X) = e^{1/2},\ \mathrm{var}(e^X) = e^2 - e$.

Chapter VI
Bivariate data and distributions

DAY 20
Bivariate statistics (part 1)

20.1 কী জিনিস?

Statistics-এর মূলেই যে আছে data সে তো বলেইছি। আর data মানেই এক বা একাধিক variable-এর বিভিন্ন value. যেমন ধরো, একটা variable হল height. একটা ক্লাসে যত ছাত্র আছে, তাদের height মেপে ফেললে পাবে একটা data set. এখানে একটাই variable, তাই এই data set-টাকে বলব **univariate** (এখানে uni অর্থ "এক")। যদি দুটো variable নিয়ে কাজ হত, তবে data set-টাকে বলতাম **bivariate.** তিনটে থাকলে trivariate বলা চলে, কিন্তু সাধারণতঃ তখন এক কথায় **multivariate** বলে ছেড়ে দিই (multi মানে "একাধিক")। যাই হোক, আমরা এখানে খালি bivariate নিয়েই আলোচনা করব। যেমন ধরো, দুটো variable হতে পারে height আর weight. মানে আমরা প্রত্যেকটা ছাত্রের উচ্চতা এবং ওজন দুটোই মাপছি। আমরা এরকম data set-কে সাধারণতঃ জোড়ায় জোড়ায় লিখে তালিকার আকারে প্রকাশ করতে পারি, যেমন

$$(h_1, w_1), ..., (h_n, w_n).$$

এখানে h_1 হল প্রথম ছাত্রের উচ্চতা, w_1 হল তার ওজন, h_2, w_2 হল যথাক্রমে দ্বিতীয় ছাত্রের উচ্চতা ও ওজন, এইরকম।

20.2 কেন?

আমরা bivariate data নিয়ে মাথা ঘামাই কেন? দুটো variable-কে আলাদা আলাদা করে দুটো univariate data set বলে ভাবলে অসুবিধা কি ছিল? অবশ্যই সেভাবে ভাবা যেত, কিন্তু তাতে একটা জিনিস করা যেত না, যেটা bivariate data set দিয়ে করা যায়--variable-দুটোর মধ্যে কোনো সম্পর্ক আছে কিনা সেটা দেখা। মনে করো আমরা জানতে চাই উচ্চতা এবং ওজনের মধ্যে সম্পর্ক আছে কিনা। যদি তোমাকে 20 জন ছাত্রের উচ্চতা এবং অন্য 20 জন ছাত্রের ওজন বলি, তাতে লাভ হবে না। এখানে এরা দুটো আলাদা univariate data set. তোমার দরকার একই ছাত্রদের উচ্চতা এবং ওজন দুটোই জানা। এখানেই bivariate data-র কেরামতি।

20.3 কত রকম?

Variable যে প্রধানতঃ দুরকম হয় সে তো জানোই, discrete আর continuous. সেই অনুযায়ী bivariate data set-ও তিনরকমের হতে পারে-- দুটো variable-ই continuous, কিংবা দুটোই discrete, কিংবা একজন discrete আর অন্যজন continuous. একটা করে উদাহরণ দেখা যাক।

Example 1: এক্ষুণি যে height, weight মাপার কথা বললাম, সেখানে দুটো variable-ই continuous. ■

Example 2: যদি তুমি একটা গ্রামে গিয়ে সেখানকার প্রতিটি প্রাপ্তবয়স্ক মানুষের কাছ থেকে এই দুটো প্রশ্নের উত্তর নাও--

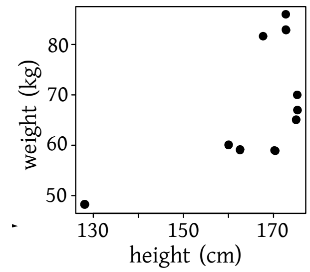


Fig 1

- আপনার ধর্ম বিশ্বাস কী? [হিন্দু, মুসলমান, খ্রীষ্টান , নাস্তিক, অন্যকিছু]
- আপনি কি সাক্ষর? [হ্যাঁ, না],

তবে তোমার কাছে দুটো variable থাকবে "ধর্মবিশ্বাস" এবং "সাক্ষরতা"। এরা দুজনেই discrete. ■

Example 3: ধূমপানের সঙ্গে রক্তচাপের কোনো সম্পর্ক আছে কি? সেটা বোঝার জন্য কিছু লোকের প্রত্যেকের রক্তচাপ মাপা হল এবং ধূমপায়ী কিনা সেটাও জেনে নেওয়া হল। এখানে দুটো variable হল "রক্তচাপ" আর "ধূমপান করে কিনা" । প্রথমটা continuous, দ্বিতীয়টা discrete. ■

আমরা এই বইতে মূলতঃ প্রথমক্ষেত্রটা নিয়েই আলোচনা করব, যেখানে দুটো variable-ই continuous. এক্ষেত্রে ওদেরকে এক ধরণের গ্রাফ এঁকে দেখানো যায়, যার নাম **scatterplot.** সেটা কী জিনিস বুঝে নেওয়া যাক।

20.4 ছবি এঁকে দেখানো

ধরো আমাদের variable দুটো হল height আর weight. একটা ক্লাসের দশটা ছাত্রের height আর weight মেপে আমরা পেয়েছি

$$(h_1, w_1), \quad \cdots, \quad (h_{10}, w_{10}).$$

এই দুটো variable-কে আমরা গ্রাফের দুটো axis-এ দেখাব। যেকোনো axis-এ যেকোনো variable দেখানো যেতে পারে, কিন্তু কোন axis-এ কোনটা দেখালে ভালো লাগবে, সে বিষয়ে একটা সহজ নিয়ম আছে। দুটো variable-এর মধ্যে একটা যদি অন্যটার কারণ হয়, তবে সেই "কারণ" variable-টাকে সাধারণতঃ দেখানো হয় horizontal দিকে। এখানে যেমন height-কে weight-এর কতকটা কারণ মনে করা যায়। কারণ একটা লম্বা লোকের ওজন বেশী হবে এটাই প্রত্যাশিত, কিন্তু ওজন বেশী হলেই লম্বা হবে এমনটা মনে করার কোনো কারণ নেই। সেই কারণে আমরা এখানে height-কে দেখাব horizontal দিকে, আর weight যাবে vertical দিকে। যদি variable দুটোর মধ্যে এরকম কার্যকারণ সম্পর্ক না থাকে, তবে যেটাকে খুশি যেদিকে দেখাতে পারো। এবার $(h_1, w_1), ..., (h_{10}, w_{10})$-রা তাহলে গ্রাফের উপরে দশটা বিন্দু। প্রতিটা বিন্দুতে আমরা একটা করে ফুটকি আঁকলেই scatterplot হয়ে গেল। Fig 1 দ্যাখো। এখানে বিন্দুগুলোকে লাইন দিয়ে যোগ করা হয় না। একটা scatterplot থেকে দুটো variable-এর মধ্যে কীরকম সম্পর্ক, সে বিষয়ে একটা ধারণা করা যায়। সেই আলোচনাই এবার করব।

Fig 2

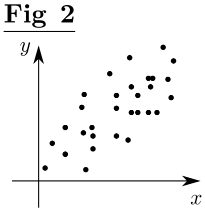


Fig 3

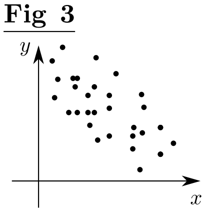


Fig 4

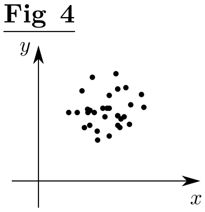
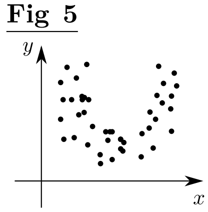


Fig 5



20.5 Correlation

Fig 2, Fig 3, Fig 4 আর Fig 5-এ চার রকমের scatterplot দেখানো হয়েছে। প্রথম ক্ষেত্রে মোটামুটিভাবে বলা যায় যে, x বাড়লে y-ও বাড়ছে। "x-এর বৃদ্ধি" মানে ডানদিকে সরে যাওয়া (বা ম্যাপের মত করে ভাবলে পশ্চিম থেকে পূর্বদিকে)। একইভাবে "y-এর বৃদ্ধি" মানে দক্ষিণ থেকে উত্তরে যাওয়া। লক্ষ কর যে, Fig 2-এ বিন্দুগুলো যে ছায়াপথের সৃষ্টি করেছে সেটা দক্ষিণ-পশ্চিম থেকে উত্তর-পূর্বে বিস্তৃত। সেই জন্যই বললাম যে x বাড়লে মোটামুটিভাবে y-ও বাড়ছে। Fig 3-এর বেলায় ব্যাপারটা ঠিক উল্টো, এখানে মোটের উপর বলা যায় যে, x বাড়লে y কমছে। Fig 4-কে বলা যায় এদের মাঝামাঝি, এখানে x আর y-এর মধ্যে খুব একটা সম্পর্ক আছে বলে মনে হচ্ছে না। একটা সম্পূর্ণ অন্যরকম পরিস্থিতি রয়েছে Fig 5-এ। এখানে x বাড়লে y প্রথমে কমছে, তারপরে বাড়ছে। আমরা এবার একটা অংকের কায়দা শিখব, যা দিয়ে প্রথম তিনটে ক্ষেত্রকে আলাদা করা যাবে, অর্থাৎ $(x_1, y_1), ..., (x_n, y_n)$ দেওয়া থাকলে scatterplot না এঁকেই কেবলমাত্র সেই কায়দাটা করেই বলে দেওয়া যাবে x বাড়লে মোটামুটিভাবে y-ও বাড়ে নাকি কমে, নাকি কোনোটাই করে না। এই কায়দাটার নাম **correlation.** তবে দুঃখের বিষয় যে, Fig 5-এর মত পরিস্থিতিতে correlation-এর কায়দাটা খাটবে না। তা না খাটুক, বাস্তবে এমন নানা উদাহরণ আছে যেখানে জানাই থাকে যে, scatterplot-টা প্রথম তিনটে ছবির কোনো একটার মত হবে। সে সব জায়গায় আমরা নিশ্চিন্তে correlation লাগাতে পারব।

Correlation-এর ধারণাটা খুব যে কঠিন এমন নয়, তবে ধাপে ধাপে বুঝলে সুবিধা হবে। প্রথম ধাপে শিখব **covariance** বলে একটা জিনিস। Correlation-এর মূল শক্তিটা আসে এই covariance থেকেই। এর জন্যে প্রথমে আমরা দেখি x-এর হ্রাস বৃদ্ধি কী করে মাপা যায়। একটা উদাহরণ নিই।

Example 4: ধরো বললাম রায়বাবু একশ দশ বছর বেঁচেছিলেন। তুমি শুনে বলবে--ওরেব্বাবা তবে তো তিনি দীর্ঘজীবী

মানুষ ছিলেন! আবার যখন শুনবে যে, কবি সুকান্ত বাইশ বছর বয়সে মারা যান, তোমার মনে--হবে আহা রে, কত অল্পবয়সেই চলে গেল। আমাদের এইরকম বলার কারণ হল এই যে, আমরা জানি ভারতীয়দের গড় আয়ু মোটামুটি কত হয়। এই ধর ষাট[1]। সুতরাং মৃত্যুর সময়ে কত বয়স ছিল সেটাকে আমরা মনে মনে এই গড় বয়সটার সঙ্গে তুলনা করি। যদি খুব বেশী হয় তবে বলি "হুম, দীর্ঘজীবী বটে"। আর যদি অনেক কম হয় তবে, ফোঁস করে দীর্ঘনিঃশ্বাস ছেড়ে বলি "আহা, অকালেই ঝরে গেল রে।" ■

এই ধারণাটা আমরা খাটাব x-এর উপরে। এখানে x-এর যে কটা value দেওয়া আছে, তারা হল $x_1, ..., x_n$. এদের গড় বার করব, মানে sample mean–

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

তারপর দেখব x_i-রা এই $\bar{x}$-এর কতটা উপরে বা নীচে, মানে এই সংখ্যাগুলো বার করব

$$x_1 - \bar{x},\ \ldots\ ,\ x_n - \bar{x}.$$

একই কাজ করব y-দের নিয়েও, ফলে পাব

$$y_1 - \bar{y},\ \ldots\ ,\ y_n - \bar{y}.$$

এবার পরীক্ষা করে দেখব x বাড়লেই y-ও বাড়ছে কিনা। তার জন্য একটা ছোট্টো কৌশল করা যায়--$(x_i - \bar{x})$আর $(y_i - \bar{y})$-দেরকে জোড়ায় জোড়ায় গুণ করে দেব, মানে $(x_i - \bar{x}) \times (y_i - \bar{y})$ বার করব। ফলে পাব এই n-টা সংখ্যা--

$$(x_1 - \bar{x})(y_1 - \bar{y}),\ \ldots\ ,\ (x_n - \bar{x})(y_n - \bar{y}).$$

আমরা জানি যে, দুটো positive সংখ্যাকে গুণ করলে যেমন positive গুণফল হয়, তেমনি দুটো negative সংখ্যাকে গুণ করলেও positive-ই হয়। সুতরাং যদি কোনো x_i বড় হয়, আর তার y_i-টাও বড় হয়, মানে $x_i - \bar{x}$ আর $y_i - \bar{y}$ দুজনেই positive হয়, তবে $(x_i - \bar{x})(y_i - \bar{y})$-ও positive হবে। আবার যদি দুজনেই negative হয়, তাহলেও $(x_i - \bar{x})(y_i - \bar{y}) > 0$ হবে। তাহলে $(x_i - \bar{x})(y_i - \bar{y}) < 0$ কখন হবে? যখন x_i, y_i-এর মধ্যে একজন বড়, কিন্তু অন্যজন ছোটো, তখন। এইভাবে আমরা দুটো দল পেলাম--

[1]সরকারী হিসেব অনুযায়ী 62.3 বছর।

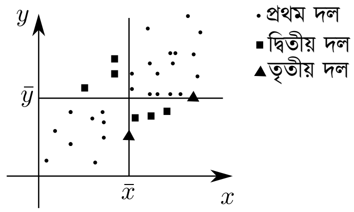


Fig 6

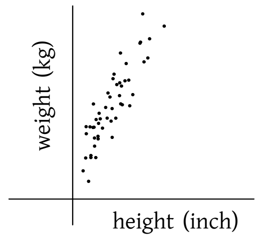


Fig 7

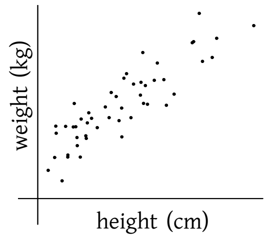


Fig 8

- একদলে আছে তারা, যাদের জন্য $(x_i - \bar{x})(y_i - \bar{y}) > 0$, অর্থাৎ x_i, y_i হয় দুজনেই একইসঙ্গে বড়, আর নয়তো একইসঙ্গে ছোটো।
- আর অন্যদলে আছে তারা, যাদের জন্য $(x_i - \bar{x})(y_i - \bar{y}) < 0$, মানে একজন ছোটো অন্যজন বড়।

অবশ্য সূক্ষ্ম বিচারে দেখলে একটা তৃতীয় দলও হওয়া উচিত, যেখানে $(x_i - \bar{x})(y_i - \bar{y}) = 0$. এখানে হয় $x_i = \bar{x}$ বা $y_i = \bar{y}$. বোঝার জন্য Fig 6 দ্যাখো। এবার দেখি এই তিনদলের মধ্যে সবচেয়ে ভারী কোন দল। সেটা বার করা জন্য এদের সবার গড় বার করব। এরই নাম হল covariance of x, y, যাকে সংক্ষেপে লেখে $\text{cov}(x, y)$. মানে

$$\text{cov}(x, y) = \frac{1}{n}\sum_{1}^{n}(x_i - \bar{x})(y_i - \bar{y}).$$

যদি এটা > 0 হয়, তবে বুঝতে হবে যে প্রথম দলটা বেশী ভারী (মানে x বাড়লে y-ও বাড়ছে)। যদি covariance-টা < 0 হয়, তবে বুঝব যে দ্বিতীয় দলটা বেশী ভারী (মানে x বাড়লে y কমছে)। যদি covariance-টা $= 0$ হয়, তবে অবশ্য কী হচ্ছে বলা মুস্কিল। এমন হতে পারে যে, তৃতীয় দলটা সবচেয়ে ভারী, বা এমনও হতে পারে যে, প্রথম দুটো দল সমান সমান ভারী, তাই কাটাকাটি হয়ে 0 হয়ে গেছে।

Covariance-এর ধারণাটা একবার বুঝে গেলে correlation-এর ধারণাটা অনায়াসেই বুঝবে, খালি একটা বাড়তি ধাপ লাগবে। সেই ধাপটা বোঝার জন্য আবার covariance-এর দিকে চোখ ফেরানো যাক। সেখানে আমরা $(x_i - \bar{x})(y_i - \bar{y})$-জাতীয় জিনিস নিয়ে কাজ করছিলাম। আমরা এতক্ষণ খালি এটুকুই দেখছিলাম যে সেটা > 0, নাকি < 0, নাকি $= 0$. লক্ষ কর যে, যদি $x_i - \bar{x}$ মস্ত বড় কিছু হয় আর $y_i - \bar{y}$-ও খুব বড় কিছু হয়, তবে $(x_i - \bar{x})(y_i - \bar{y})$ শুধু > 0 হবে তাই নয়, খুব বড় একটা positive সংখ্যা হবে। ফলে covariance-এ এর অবদান হবে অনেকটা। যদি $(x_i - \bar{x})(y_i - \bar{y})$ একটা ছোটো positive সংখ্যা হত, তাহলেও একটা অবদান থাকত, কিন্তু অপেক্ষাকৃত কম। এই কথাটা মাথায় রেখে নীচের উদাহরণটা দ্যাখো।

Example 5: একটা ক্লাসে অনেক ছাত্র আছে। তাদের ওজন আর উচ্চতা মেপে আমরা দুটো scatterplot পেয়েছি,

যেগুলোকে দেখিয়েছি Fig 7 আর Fig 8-এ। প্রথম ক্ষেত্রে উচ্চতা মেপেছি ইঞ্চিতে। দ্বিতীয়ক্ষেত্রে সেন্টিমিটারে। ওজন দুইক্ষেত্রেই কেজিতে মেপেছি। তাই দুটো ছবিই আসলে একই, তফাত কেবল এককে। যদি covariance বার করি তবে, দ্বিতীয়ক্ষেত্রে covariance-টা 2.54 গুণ বড় আসবে, কারণ এখানে উচ্চতাগুলো (মানে x_i-গুলো) 2.54 দিয়ে গুণ হয়ে গেছে। কিন্তু তার মানে কি এই যে, উচ্চতা আর ওজনের সম্পর্ক এককের উপর নির্ভর করে? সেটা তো হতে পারে না, কারণ ওজন আর উচ্চতার মধ্যে সম্পর্ক হল প্রাকৃতিক ব্যাপার, কিন্তু একক হল মানুষের মনগড়া ধারণা। সেই কারণে covariance-কে বিচার করার সময়ে আমরা x আর y-এর একক দুটোকে হিসেব থেকে বাদ দিতে চাই। এর একটা পথ হল covariance-কে ওদের দুজনেরই standard deviation দিয়ে ভাগ করে দেওয়া। এর ফলে যেটা পাবে তাকেই বলে **correlation.** ■

পুরো ব্যাপারটাকে গুছিয়ে লিখে নিই--

Example 6: Define the sample correlation coefficient r for the sample

$$(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$$

from a bivariate population. **(2014)**
SOLUTION:

DEFINITION: (Sample) covariance, correlation

Let $(x_1, y_1), ..., (x_n, y_n)$ be a bivariate data set where $n \geq 2$. Then the <u>covariance</u> between the variables $x$and $y$is defined as

$$\text{cov}(x, y) = \frac{1}{n}\sum_1^n (x_i - \bar{x})(y_i - \bar{y}).$$

The <u>correlation</u> between them is defined as

$$\text{corr}(x, y) = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x)\text{var}(y)}},$$

if $\text{var}(x), \text{var}(y) > 0$. If at least one of $\text{var}(x), \text{var}(y)$ is 0, then $\text{corr}(x, y)$is undefined.

The correlation is often denoted by r.

■

Covariance বার করার একটা সহজতর পথ আছে, এইরকম--

THEOREM

$$\text{cov}(x, y) = \frac{1}{n}\sum_1^n x_i y_i - \bar{x}\bar{y}.$$

এবার একটা অংক করি, যাতে ফর্মুলাগুলো সব সড়গড় হয়ে যায়।

Example 7: A bivariate sample of size 11 gave the results $\bar{X} = 7$, $S_x = 2$, $\bar{Y} = 9$, $S_y = 4$, and $r = 0.5$. It is later found that one pair of sample values $(X = 7, Y = 9)$ was inaccurate and was rejected. How would the original value of r be affected by the rejection?[4] **(2010.4c)**
SOLUTION: এখানে S_x মানে X-এর standard deviation. তাই

$$S_x^2 = \frac{1}{n}\sum X_i^2 - (\bar{X})^2. \qquad (*)$$

একইভাবে S_y হল Y-এর standard deviation. এখানে r মানে হচ্ছে $\text{corr}(x, y)$.
এই অংকটা দুইভাবে করা যায়। এক হল $\sum X_i, \sum X_i^2$ ইত্যাদি বার করে। প্রথমে সেভাবেই এগোই।
ভুলটা ধরা পড়ার আগে হিসেবটা কীরকম ছিল, দেখি--

Originally,

$$
\begin{aligned}
n &= 11,\\
\sum X_i &= n\bar{X} = 11 \times 7 = 77,\\
\sum Y_i &= n\bar{Y} = 11 \times 9 = 99,\\
\sum X_i^2 &= n(S_x^2 + (\bar{X})^2) = 11(2^2 + 7^2) = 583,\\
&\quad \text{এটা পাচ্ছি } S_x^2\text{-এর সংজ্ঞা থেকে। (*) দেখে নিতে পারো।}\\
\sum Y_i^2 &= n(S_y^2 + (\bar{Y})^2) = 11(4^2 + 9^2) = 1067,\\
\sum X_iY_i &= n(rS_xS_y + \bar{X}\bar{Y}) = 11(0.5 \times 4 \times 2 + 7 \times 9) = 737.\\
&\quad \text{এখানে আমরা ব্যবহার করলাম } \operatorname{cov}(X,Y) = rS_xS_y.
\end{aligned}
$$

এইবার দেখি হিসেব সংশোধন করা পর (মানে $(X = 7, Y = 9)$ কেসটাকে বাদ দেওয়ার পর) অবস্থাটা কী দাঁড়িয়েছে--

After removing the $(X = 7, Y = 9)$ case,

$$
\begin{aligned}
n &= 11 - 1 = 10,\\
\sum X_i &= 77 - 7 = 70,\\
\sum Y_i &= 99 - 9 = 90,\\
\sum X_i^2 &= 583 - 7^2 = 534,\\
\sum Y_i^2 &= 1067 - 9^2 = 986,\\
\sum X_iY_i &= 737 - 7 \times 9 = 674.
\end{aligned}
$$

So

$$
\begin{aligned}
\bar{X} &= \frac{1}{n}\sum X_i = \frac{70}{10} = 7,\\
\bar{Y} &= \frac{1}{n}\sum Y_i = \frac{90}{10} = 9,\\
S_x^2 &= \frac{1}{n}\sum X_i^2 - (\bar{X})^2 = \frac{534}{10} - 7^2 = 4.4,\\
S_y^2 &= \frac{1}{n}\sum Y_i^2 - (\bar{Y})^2 = \frac{986}{10} - 9^2 = 17.6,\\
S_{xy} &= \frac{1}{n}\sum X_iY_i - (\bar{X})(\bar{Y}) = \frac{674}{10} - 7 \times 9 = 4.4.
\end{aligned}
$$

So the new value of r is

$$
r = \frac{S_{xy}}{S_xS_y} = \frac{4.4}{\sqrt{4.4 \times 17.6}} = 0.5.
$$

> So r does not change.

■

Exercise 1: এতক্ষণ আমরা প্রথম পদ্ধতিতে অংকটা করলাম। এতে বেশ কসরত না করতে হল। কিন্তু এতসব না করেই এক্ষেত্রে বোঝা সম্ভব ছিল যে r অপরিবর্তিত থাকবে। সেটাই হল দ্বিতীয় পদ্ধতি। কী করে বলো তো?
■

20.6 Covariance আর correlation-এর কিছু ধর্ম

এবার আমরা covariance আর correlation-এর কিছু ধর্ম শিখব, যেগুলো সরাসরি সংজ্ঞা ব্যবহার করলেই দু-এক লাইনে প্রমাণ করা যায়।

THEOREM

Let $(x_1, y_1), ..., (x_n, y_n)$ be a bivariate data set. Let $a, b, c, d \in \mathbb{R}$ be any numbers, where $a, c \neq 0$. Then

$$\text{cov}(ax + b, cy + d) = ac\,\text{cov}(x, y).$$

অর্থাৎ কিনা, গুণ হওয়া সংখ্যারা covariance-এর বাইরে বেরিয়ে আসে, আর যোগ হওয়া সংখ্যারা উড়ে যায়।
এটা থেকে correlation-এর একটা ধর্ম পাওয়া যায়--

THEOREM

Let $(x_1, y_1), ..., (x_n, y_n)$ be a bivariate data set. Let $a, b, c, d \in \mathbb{R}$ be any numbers, where $a, c \neq 0$. Then

$$\text{corr}(ax + b, cy + d) = sign(ac)\text{corr}(x, y).$$

যদি $a_1, ..., a_n$ আর $b_1, ..., b_n$ কিছু real number হয়, তবে

$$\left(\sum a_i b_i\right)^2 \leq \sum a_i^2 \sum b_i^2$$

হতে বাধ্য। একে বলে Cauchy-Schwartz inequality. যদি $a_i = x_i - \bar{x}$আর $b_i = y_i - \bar{y}$ নাও, তবে অমনি পাবে--

THEOREM

Let $(x_1, y_1), ..., (x_n, y_n)$ be a bivariate data set. Then

$$-1 \leq \text{corr}(x, y) \leq 1.$$

এবার covariance-এর কিছু ধর্মের কথা বলব, যেগুলোর মূল কথা হল "covariance ব্যাপারটা গুণ করার মত আচরণ করে"। যেমন ধরো $xy = yx$ হয়। একইভাবে $\text{cov}(x,y) = \text{cov}(y,x)$ হয়। আবার $x \times x = x^2$ হয়। এখানে square-এর ভূমিকা পালন করে *variance*. তাই $\text{cov}(x,x) = \text{var}(x)$ হয়। আমরা জানি যে $(x+y)z = xz + yz$ হয়। Covariance-এর বেলাতেও $\text{cov}(x+y,z) = \text{cov}(x,z) + \text{cov}(y,z)$ হয়। আরেকটা উদাহরণ দেখাই--আমরা জানি $(ax+by)(cx+dy) = acx^2 + (ad+bc)xy + bdy^2$. ঠিক একইরকম ব্যাপার হবে covariance-এর ক্ষেত্রেও-- $\text{cov}(ax+by, cx+dy) = ac\text{var}(x) + (ad+bc)\text{cov}(x,y) + bd\text{var}(y)$.

THEOREM

Let $(x_1,y_1),...,(x_n,y_n)$ be a bivariate data set. Let $a,b,c,d \in \mathbb{R}$ be fixed numbers. Then

1. $\text{cov}(x,y) = \text{cov}(y,x)$
2. $\text{cov}(x,x) = \text{var}(x)$
3. $\text{cov}(ax+by, cx+dy) = ac\text{var}(x) + (ad+bc)\text{cov}(x,y) + bd\text{var}(y)$.

এদের ব্যবহার করে ছোটোবেলায় শেখা বিভিন্ন ফর্মুলার covariance-সংস্করণ পাওয়া যায়। যেমন $(a+b)^2 = a^2+2ab+b^2$ থেকে পাব

$$\text{var}(x+y) = \text{var}(x) + 2\text{cov}(x,y) + \text{var}(y).$$

আবার $(a-b)(a+b) = a^2 - b^2$ থেকে আসবে

$$\text{cov}(x+y, x-y) = \text{var}(x) - \text{var}(y).$$

ব্যাপারটা খালি দুটো variable-এর ক্ষেত্রেই সীমাবদ্ধ নয়। যেকোনো finite-সংখ্যক variable-এর ক্ষেত্রেও খাটে। যেমন ধরো, আমরা জানি যে

$$\left(\sum_{i=1}^n a_i\right)^2 = \sum_i a_i^2 + 2\sum_{i<j} a_i a_j.$$

এর covariance-সংস্করণ হল

$$\text{var}\left(\sum_{i=1}^n x_i\right) = \sum_i \text{var}(x_i) + 2\sum_{i<j} \text{cov}(x_i, x_j).$$

DAY 21
Bivariate statistics (part 2)

21.1 Least square method

ধরো তোমাকে Fig 9-এর মত একটা scatterplot দিলাম। লক্ষ কর যে, point-গুলো মোটামুটিভাবে একটা সরলরেখা বরাবর আছে। তোমাকে যদি চোখের আন্দাজে এই সরলরেখাটা আঁকতে বলি, তবে তুমি হয়তো Fig 10-এর মত আঁকবে। এই কাজটা অনেক সময়েই করার দরকার পড়ে। কিন্তু কেবলমাত্র চোখের আন্দাজে এগোতে গেলে দুটো সমস্যা হয়--

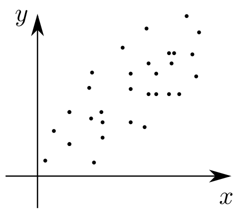


Fig 9

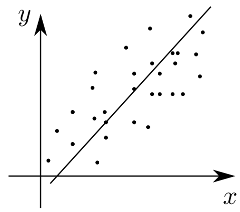


Fig 10

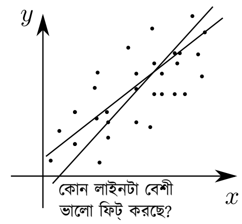


Fig 11

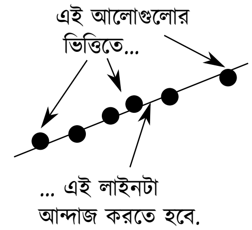


Fig 12

- এক, চোখের আন্দাজ হল যাকে বলে subjective, অর্থাৎ তোমার চোখে যেটাকে ভালো মনে হচ্ছে, সেটা আমার কাছে ভালো নাই মনে হতে পারে। যেমন Fig 11-এ যে দুটো লাইন আছে, তাদের মধ্যে কোনটা বেশী ভালো ফিট করছে, সেটা চোখের আন্দাজে ঠিক করা কঠিন।

- আর দুই, সবসময়ে যে খালি সরলরেখাই আঁকতে হবে এমন নয়, কখনও parabola বা hyperbola বা আরো জটিল কোনো কিছুও আঁকতে হতে পারে, তখন চোখের আন্দাজ মোটেই কাজ আসবে না।

এই কারণে আমরা একটা অংকের কায়দা বার করে রাখতে চাই যেটা চোখের আন্দাজের কাজটা করে দেবে। তাতে subjective হয়ে যাওয়ার ভয়ও থাকবে না, আর জটিলতর ক্ষেত্রেও কাজ করা যাবে। আরেকটা সুবিধাও আছে। সেটা artificial intelligence-এ কাজে লাগে। ধরো একটা রোবট একটা সোজা লাইন বরাবর যাবে। সেই লাইনটা বোঝানোর জন্য কিছু দূর অন্তর অন্তর একটা করে উজ্জ্বল আলো বসানো আছে Fig 12-র মত। তুমি চোখে দেখেই বুঝে যাচ্ছ যে এরা মোটামুটি একটা লাইন বরাবর আছে। কিন্তু রোবটের চোখ তো একটা ক্যামেরামাত্র, তার তো আর কল্পনাশক্তি নেই। সুতরাং তার কম্পিউটার মগজে প্রোগ্রাম করে অংক ঢুকিয়ে দিতে হয়, যাতে ওই বিচ্ছিন্ন কয়েকটা আলোর বিন্দু থেকে লাইনটা বুঝে নিতে পারে। আমরা যে কায়দাটা শিখতে চলেছি, সেই কায়দায় এই artificial intelligence-এর সমস্যারও সমাধান করা যায়।

21.1.1 অংকের ভাষায়

ব্যাপারটা অংকের ভাষায় লেখার জন্য হায়ার সেকণ্ডারীর physics থেকে একটা উদাহরণ নিই। তোমরা জানো যে একটা স্প্রিঙের থেকে ওজন ঝোলালে তার দৈর্ঘ্য বেড়ে যায় (Fig 13)। এই নিয়ে একটা সূত্র আছে যে, লম্বায় কতটা বাড়বে সেটা ঝোলানো ওজনের সঙ্গে proportional (সমানুপাতিক) হবে। মানে স্প্রিংটার স্বাভাবিক দৈর্ঘ্য যদি a হয় আর x পরিমাণ ওজন ঝোলালে যদি দৈর্ঘ্য বেড়ে y হয়, তবে $y-a\propto x$ হবে মানে $y-a=bx$, যেখানে b হল একটা constant, যেটা স্প্রিংটার নিজস্ব বৈশিষ্ট্য। এটাকে সাজিয়ে এভাবে লেখা যায়--

$$y=a+bx,$$

যেটা একটা সরলরেখার equation. এবার ধরো একটা স্প্রিঙে $x_1,...,x_n$ ওজনের বাটখারা চাপিয়ে প্রতিক্ষেত্রে দৈর্ঘ্য মেপে দেখা গেল যথাক্রমে $y_1,...,y_n$ হয়েছে। প্রশ্ন হল--এই স্প্রিংটার জন্য a আর b কী কী? যদি গ্রাফে আঁকি তবে Fig 14-এর মত কিছু একটা হবে। দেখাই যাচ্ছে যে এই point-গুলো মোটেই একই সরলরেখার উপরে নেই, যদিও কাছাকাছি আছে। একেবারে নিখুঁতভাবে নেই কেন? তাহলে কি আমাদের স্প্রিংটার বেলায় সূত্রটা খাটে না? না, আসলে আমরা যখন দৈর্ঘ্যটা মাপছি, তখন সেটা নিখুঁত হচ্ছে না। যদি হত তবে আমরা সত্যিই একটা সরলরেখা পেতাম। একটু বুঝে নিই যে, এই ভুলটা

Fig 13

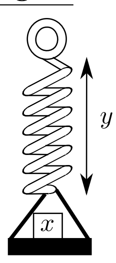


Fig 14

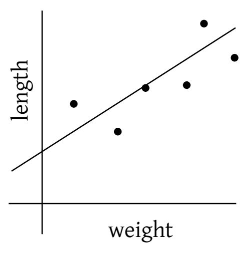


Fig 15

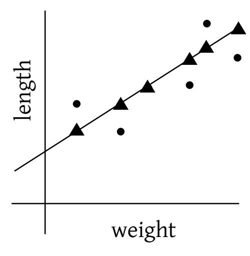
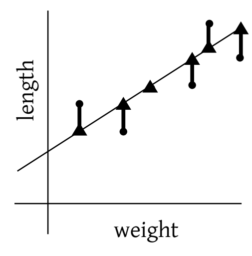


Fig 16



কীরকম। লক্ষ করো যে, $x = x_i$-এর জন্য নির্ভুল মাপটা হওয়া উচিত ছিল $a + bx_i$. কিন্তু মেপে পেয়েছি y_i, যেটা মাপার সামান্য এদিক ওদিকের জন্য একেবারে $a + bx_i$-এর সমান নাও হতে পারে। তার মানে গ্রাফের উপরে $(x_i, a + bx_i)$-তে একটা বিন্দু থাকার কথা ছিল, তার জায়গায় বিন্দুটা গিয়ে পড়েছে (x_i, y_i)-তে। যেহেতু দুইক্ষেত্রেই x_i একই আছে, তাই দুটো বিন্দুই রয়েছে একই vertical লাইনে। তার কারণ বাটখারাটা তো আর আমরা মাপছিনা, ওটা কারখানা থেকে একদম তৈরী হয়েই এসেছে। তাই ওটা x_i-টা সবসময়েই x_i-ই থাকছে। মাপতে গিয়ে ভুলচুক যাই হোক না কেন, তাতে x-টা বদলাবে না, বদলে যেতে পারে y. মাপার কোনো ভুল না থাকলে যে $(x_i, a + bx_i)$-জাতীয় point-গুলো পেতাম, তারা সবাই আছে $y = a + bx$ সরলরেখার উপরে, Fig 15-এ ত্রিভুজ দিয়ে দেখানো বিন্দুগুলোর মত। সুতরাং কোনো গোল বিন্দুর নিখুঁত সংস্করণটা হবে লাইনের উপরে সেই ত্রিভুজ বিন্দুটা, যেটা ওর সঙ্গে একই vertical লাইনে আছে। তার মানে মোট ভুলের পরিমাণ মাপা যায় Fig 16-এর মোটা লাইনগুলোর দৈর্ঘ্য থেকে, মানে $y_i - (a + bx_i)$-গুলোর থেকে। এখানে মনে রেখো যে ভুলটা বেশীর দিক না কমের দিকে তা নিয়ে আমাদের মাথা ব্যথা নেই। আমরা $y_i - (a + bx_i)$-গুলোকে square করে যোগ করব, এবং যোগফলটাকে বলব মোট ভুলের পরিমাণ, যার নাম দিলাম ধরো $S(a, b)$, মানে--

$$S(a,b) = \sum_1^n (y_i - (a + bx_i))^2.$$

আমরা যখন মাপছিলাম, তখন ভুলচুক একটুআধটু এড়ানো যায়নি বটে, কিন্তু নিশ্চয়ই আমরা সচেষ্ট ছিলাম যাতে ভুলটা কম হয়। সেই কারণে $S(a, b)$ বেশ ছোটো হওয়াই প্রত্যাশিত। তাই আমরা এমন a আর b খুঁজব যাতে $S(a, b)$-টা minimum হয়। এই কাজটা করার জন্য আমরা একটা অংকের কায়দা ব্যবহার করব। যদি কোনো $f(x)$-এর minimum বার করতে হয়, তবে তার একটা কায়দা হল $f'(x) = 0$ করা, এবং $f''(x) > 0$ হচ্ছে কিনা দেখা। এখানে $S(a, b)$ হল a আর b দুজনের function,তাই এখানে আমরা প্রথমে $\frac{\partial S}{\partial a}$ এবং $\frac{\partial S}{\partial b}$ দুজনকেই 0-র সমান করব--

$$\begin{aligned} \frac{\partial S}{\partial a} &= -2\sum_1^n (y_i - a - bx_i) = 0, \\ \frac{\partial S}{\partial b} &= -2\sum_1^n (y_i - a - bx_i)x_i = 0. \end{aligned}$$

এদের থেকে এই দুটো equation পাবে, যাদের বলে **normal equations**–

$$\begin{aligned} na + \left(\sum x_i\right) b &= \sum y_i, \\ \left(\sum x_i\right) a + \left(\sum x_i^2\right) b &= \sum x_i y_i. \end{aligned}$$

এবার এদের থেকে a আর b বার করতে হবে। এভাবে যে উত্তর পেলাম তার নাম দিলাম ধরো $\widehat{a}$ আর $\widehat{b}$. এই $\widehat{a}, \widehat{b}$-কেই বলব a, b-এর **least square estimator.**

এবার আমরা $S(a, b)$-র second derivative বার করব। এখানে যেহেতু partial derivative নিয়ে কাজ করছি, তাই second derivative-টা এখানে একটা matrix, যার পোশাকি নাম হল **Hessian matrix.** আমরা এখানে a, b-র সাপেক্ষে differentiate করছি, তাই এই matrix-টা হবে

$$H(a,b) = \begin{bmatrix} \frac{\partial^2 S}{\partial a^2} & \frac{\partial^2 S}{\partial a \partial b} \\ \frac{\partial^2 S}{\partial a \partial b} & \frac{\partial^2 S}{\partial b^2} \end{bmatrix}.$$

এখানে দেখতে হবে $H(\widehat{a}, \widehat{b})$-টা positive definite হচ্ছে কিনা, অর্থাৎ--

- এটা nonsingular কিনা,
- এবং একে $A'A$ আকারে লেখা যায় কিনা, যেখানে A হল কোনো একটা matrix, আর A' তার transpose.

আমাদের উদাহরণে দেখবে Hessian-টা আসছে এইরকম--

$$H = 2\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}.$$

এটা nonsingular কিনা সেটা দেখার জন্য H-এর determinantবার করো--

$$det(H) = n\sum x_i^2 - \left(\sum x_i\right)^2 = n^2\text{var}(x).$$

যদি x_i-গুলো সব সমান না হয়, তবে $\text{var}(x) > 0$, তাই সেক্ষেত্রে H হবে nonsingular. সুতরাং প্রথম শর্ত মিলল। দ্বিতীয় শর্তের জন্য এইটা লক্ষ কর যে, H-এর এইভাবে লেখা যায় $H = A'A$, যেখানে

$$A = \sqrt{2}\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

(চট্ করে $A'A$ বার করেই দ্যাখো, H পাও কিনা!)
এই পুরো আলোচনার নির্যাসটুকু লিখতে বলেছে নীচের অংকটার উত্তরে।

Example 8: Explain the principle of least squares regarding fitting mathematical curves.[7] **(2006.4b)**

SOLUTION: এইটা একটা গল্প-লিখতে-বলা প্রশ্ন। এরকম ক্ষেত্রে একটা উদাহরণ মাথায় না রাখলে গল্পটার খেই হারিয়ে যেতে পারে। তাই আমরা গল্পের পাশাপাশি আমাদের স্প্রিঙের উদাহরণটা দিয়ে বোঝাতে থাকব।

> Set up: Suppose that we have a data set $(x_1, y_1), ..., (x_n, y_n)$, and a family of curves $y = f(x, \theta_1, ..., \theta_p)$ where $\theta_1, ..., \theta_p$ are unknown parameters.

আমাদের স্প্রিঙের গল্পে x_i, y_i-গুলো ছিল ওজন আর দৈর্ঘ্য। আমরা একটা সরলরেখা ফিট করার চেষ্টা করছিলাম। যাবতীয় সরলরেখার set-টা ছিল আমাদের family. এর যেকোনো সদস্যকে আমরা $y = a + bx$ আকারে লিখেছিলাম। মানে a, b ছিল আমাদের parameter, যাদের এখানে আমরা $\theta_1, ..., \theta_p$ বলেছি। অর্থাৎ স্প্রিঙের বেলায় $p = 2, \theta_1 = a$ আর $\theta_2 = b$. এখানে a, b-এর মধ্যে কোনটাকে θ_1 আর কোনটাকে θ_2 বলছি, সেটা গুরুত্বপূর্ণ নয়।
এবার লিখে নিই আমরা কী করতে চলেছি।

> Aim: We want to find values of $\theta_1, ..., \theta_p$ such that the curve $y = f(x, \theta_1, ..., \theta_p)$ passes ``as closely as possible" to the data points.

এবার বলব least square ব্যাপারটা কী।

> Least square method: For any $\theta_1, ..., \theta_p \in \mathbb{R}$ we define the sum of squared errors as
>
> $$S(\theta_1, ..., \theta_p) = \sum_1^n (y_i - f(x_i, \theta_1, ..., \theta_p))^2.$$
>
> We try to find out the values of $\theta_1, ..., \theta_p$ for which $S(\theta_1, ..., \theta_p)$ is minimised. In general such values may not exist, or may not be unique.

এইটুকু হচ্ছে least square method-এর মূল কাজটা। এই minimum বার করাটা নানাভাবে করা যেতে পারে। তার একটা হল differentiation করে। সেটা এবার বলব--

> If $S(\theta_1,...,\theta_p)$ is differentiable, then we can try to equate the partial derivatives of $S(\theta_1,...,\theta_p)$ to 0 to obtain the normal equations
>
> $$\frac{\partial S}{\partial \theta_i} = 0, \quad i = 1,...,p.$$

আমাদের স্প্রিঙের উদাহরণে $p = 2$ ছিল, তাই দুটো equation পেয়েছিলাম। সেখানে এটাও বলেছিলাম যে, solution সব সময়েই থাকবে। দুঃখের কথা এই গ্যারান্টিটা সেখানে দেওয়া গেলেও in general দেওয়া যায় না। সেই কথাটা স্বীকার করে রাখি।

> These equations may or may not have any solution in general.

যদি solution থাকে, তবে Hessian দিয়ে তার চরিত্র বিশ্লেষণ করতে হবে।

> If we can find a solution $\widehat{\theta}_1,...,\widehat{\theta}_p$ to this system, then we compute the Hessian matrix $H(\theta_1,...,\theta_p)$ which is the $p \times p$ matrix with (i,j)-th entry equal to
>
> $$\frac{\partial^2 S}{\partial \theta_i \partial \theta_j}.$$
>
> If $H(\widehat{\theta}_1,...,\widehat{\theta}_p)$ is a positive definite matrix, then $S(\theta_1,...,\theta_p)$ has a (local) minimum at $\widehat{\theta}_1,...,\widehat{\theta}_p$.
>
> In this case $\widehat{\theta}_1,...,\widehat{\theta}_p$ are taken as least square estimators of $\theta_1,...,\theta_p$.

আমরা এক্ষুণি বললাম, যে normal equation-গুলোর solution সব সময়ে নাও থাকতে পারে। কিন্তু স্প্রিঙের অংকটায় সেটা গ্যারান্টি দেওয়া যাচ্ছিল। শুধু স্প্রিঙের অংকতেই নয়, একটা বৃহত্তর ক্ষেত্রেও এই গ্যারান্টিটা দেওয়া যায়। সেই ক্ষেত্রটার উল্লেখ করে উত্তরটাকে শেষ করি--

> In the special case where $f(x,\theta_1,...,\theta_p)$ is linear in $\theta_1,...,\theta_p$, the normal equations will always have a solution, that solution will be a global minimiser.

এখানে linear in $\theta_1,...,\theta_p$ মানে হল $f(x,\theta_1,...,\theta_p)$-এর চেহারা হবে এইজাতীয়--

কিছু একটা $\times\theta_1 + \cdots +$ কিছু একটা $\times\theta_p$.

কোথাও কোনো θ_i^2 বা $\theta_i\theta_j$ জাতীয় কিছু থাকলে চলবে না। আমাদের স্প্রিঙের অংকটা এই আকারের ছিল বলেই গ্যারান্টিটা দেওয়া গিয়েছিল। ■

এবার আরেকটা উদাহরণ দেখব, যেখানেও এই গ্যারান্টিটা দেওয়া সম্ভব।

Example 9: Explain the method of least squares to fit a second degree parabola with respect to the observed data $(x_1,y_1),...,(x_n,y_n)$.[6] **(2008.5a)**

SOLUTION: ব্যাপারটা কী হচ্ছে বোঝার জন্য Fig 17 দ্যাখো। এখানে কিছু point দেওয়া আছে (x_i,y_i). কিন্তু এখানে ওরা মোটামুটি একটা সরলরেখা বরাবর নেই, আছে একটা parabola-র মত আকারে। এরকম একটা parabola এঁকে দেখিয়েছি Fig 17-এ। আমাদেরকে সেই parabola-টা বার করতে হবে, যেটা সবচেয়ে ভালো ফিট করবে। এরকম যেকোনো parabola-র equation-ই হল $y = a + bx + cx^2$ জাতীয়। তার মানে আমাদের কাজ হল উপযুক্ত a,b,c নির্বাচন করা।

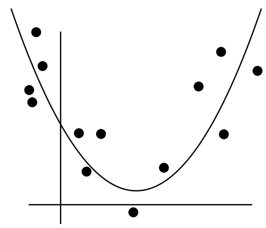


Fig 17

Suppose that we want to fit the parabola

$$y = a + bx + cx^2$$

to the data, where $a, b, c \in \mathbb{R}$ are to be estimated.
Then for each x_i the fitted value of y is $a + bx_i + cx_i^2$. So the total squared error is

$$S(a,b,c) = \sum_1^n (y_i - a - bx_i - cx_i^2)^2.$$

We want to choose $a, b, c \in \mathbb{R}$ such that $S(a,b,c)$ is the minimum possible. For this we equate the partial derivatives of S to zero:

$$\begin{aligned} \tfrac{\partial S}{\partial a} &= -2\sum_1^n (y_i - a - bx_i - cx_i^2) = 0, \\ \tfrac{\partial S}{\partial b} &= -2\sum_1^n (y_i - a - bx_i - cx_i^2)x_i = 0, \\ \tfrac{\partial S}{\partial c} &= -2\sum_1^n (y_i - a - bx_i - cx_i^2)x_i^2 = 0. \end{aligned}$$

So we get the following normal equations:

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$

এইটা ভালো করে বুঝে নাও। Matrix দিয়ে লিখেছি বলে গোড়ায় বুঝতে একটু অসুবিধা হতে পারে। যাই হোক, বুঝতেই পারছ যে, এইটাকে solve করলে যে উত্তরটা পাবে, সেটা বেশ বদখত দেখতে হবে। তাই আমরা সেটা লিখে ফেলার চেষ্টা করব না। খালি এটুকু বলে ছেড়ে দেব যে, উত্তর বার করা যায়।

It can be shown that these equations always have a solution, and that any solution minimises S.

■

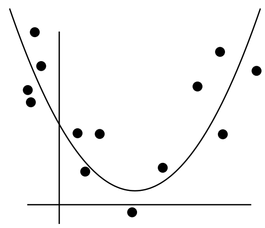


Fig 18

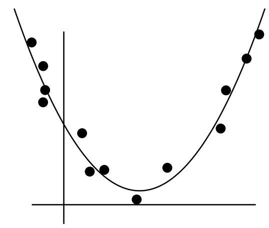


Fig 19

Example 10: Describe the method of curve fitting to a bivariate data by least square method.

Establish the normal equations for a parabolic curve to a given set of data (x_i, y_i) $(i = 1, 2, ..., n)$. How can one measure the goodness of fit?[6+4] **(2013.6a)**

SOLUTION: প্রথম দুটো অংশ তো আগেই করলাম। এবার তৃতীয় অংশ। ব্যাপারটা বুঝে নিই প্রথমে। Fig 18 আর Fig 19 দ্যাখো। দুইক্ষেত্রেই কিছু point-এ একটা করে parabola ফিট করানো হয়েছে। দ্বিতীয় ক্ষেত্রে ফিটটা খুবই ভালো, কারণ point-গুলো parabola-টার খুবই কাছাকাছি আছে। প্রথম ক্ষেত্রে বেশ কিছুটা দূর পর্যন্ত ছড়িয়ে আছে। এই তারতম্যটাকে আমরা একটা সংখ্যা দিয়ে প্রকাশ করতে চাই।

> Third part: Let the fitted values be $\widehat{y}_i$.
>
> We shall say that the fit is very good, if the correlation coefficient between y_i's and $\widehat{y}_i$'s are high.
>
> The square of this correlation is called the R^2 for this fit. It always lies in $[0, 1]$. The closer it is to 0 the worse is the fit. If it is exactly 1, then we have a perfect fit.

■

21.1.2 Straight line

আমরা এতক্ষণ least square method-এর কয়েকটা উদাহরণ দেখলাম। প্রতিক্ষেত্রেই আমরা normal equation অবধি বার করে এই বলে ছেড়ে দিচ্ছিলাম যে, এবার এগুলোকে solve করতে হবে। Solution-টা কোথাওই করে দেখাচ্ছিলাম না। তার কারণ solution-গুলোকে ক্যালকুলেটর দিয়ে বার করা কঠিন না হলেও, ফর্মুলা হিসেবে লিখলে বেশ বিচ্ছিরি দেখায়। এর একটা ব্যতিক্রম আছে। যদি আমরা খালি একটা straight line ফিট করি (যেমন স্প্রিঙের উদাহরণটার বেলায় করেছিলাম), তবে কিন্তু ফর্মুলাগুলো তেমন কঠিন নয়। করেই দেখি।

যদি আমাদের point-গুলো হয় $(x_1, y_1), ..., (x_n, y_n)$, আর তাতে আমরা $y = a + bx$ ফিট করতে চাই, তবে normal equation দুটো ছিল

$$\begin{aligned} na + \left(\sum x_i\right) b &= \sum y_i, \\ \left(\sum x_i\right) a + \left(\sum x_i^2\right) b &= \sum x_i y_i. \end{aligned}$$

হাজার হোক, মোটে দুটো equation বই তো আর কিছু নয়! সুতরাং দাঁতে দাঁত চেপে যদি solve করে $a = \widehat{a}$আর $b = \widehat{b}$ বার করেই ফ্যালো, তবে দেখবে যে,

$$\widehat{b} = \frac{\text{cov}(x, y)}{\text{var}(x)} \text{ আর } \widehat{a} = \bar{y} - \widehat{b}\bar{x}.$$

সুতরাং $y = \widehat{a} + \widehat{b}x$-কে এইভাবে লেখা যায়--

$$y - \bar{y} = \frac{\text{cov}(x,y)}{\text{var}(x)}(x - \bar{x}). \tag{**}$$

এখানে আমার কথা বিশ্বাস করেই যেন ছেড়ে দিও না, নিজের হাতে normal equation দুটো solve করে মিলিয়ে নিতে ভুলো না! এই যে straight line-টা পেলে একে বলে **least square regression line of y on x**, অর্থাৎ x-এর কোনো value দেওয়া থাকলে তা থেকে y-এর value বার করার জন্য এটাকে ব্যবহার করতে হবে। এই লাইনটার slope হল $\widehat{b}$. একে বলে **regression coefficient of y on x**. অনেক সময়ে একটা লেখে $b_{y|x}$.
যদি আমরা x আর y-এর স্থানবিনিময় করি, মানে y দেওয়া থাকে নির্ভুলভাবে, আর আমি তা থেকে x-এর value-টা আন্দাজ করার চেষ্টা করি, তবে একইভাবে পাবে **least square regression line of x on y**. ফর্মুলাটা পাওয়া যাবে আগের লাইনটার ফর্মুলায় x, y-এর স্থান বিনিময় করে দিলেই--

$$x - \bar{x} = \frac{\text{cov}(x,y)}{\text{var}(y)}(y - \bar{y}).$$

এখানে slope-টা হল $\frac{\text{cov}(x,y)}{\text{var}(y)}$, যেটাকে বলব regression coefficient of x on y. একে লেখে $b_{x|y}$.
এই দুটো ফর্মুলা মনে রাখলে ওদের মিলিয়ে মিশিয়ে নানারকম জিনিস বার করা যায়। নীচের theorem-এ তাদের কিছু পরিচয় পাবে--

THEOREM

1. $b_{x|y} \times b_{y|x} = \big(\text{corr}(x,y)\big)^2$.
2. $b_{x|y}, b_{y|x}, \text{corr}(x,y)$ and $\text{cov}(x,y)$ have the same sign.
3. The two regression lines are identical if and only if $\text{corr}(x,y) = -1$ or $\text{corr}(x,y) = 1$.
4. The two regression lines intersect at $(\bar{x}, \bar{y})$ if $\text{corr}(x,y) \in (-1, 1)$.

এগুলো প্রমাণ করা খুবই সহজ, তোমাকেই তার দায়িত্ব দিলাম।
এবার একটা প্রয়োগ দেখা যাক।

Example 11: The least square regression line of y on x and of x on y are, respectively, $x + 3y = 0$ and $3x + 2y = 0$. If the standard deviation $\sigma_x = 1$, find the least square line of v on u where $u = x + y$ and $v = x - y$.[3] **(2014.6b)**

SOLUTION: এখানে least square (regression) line of v on u বার করতে বলেছে। এরকম লাইনের চেহারাটা (**) থেকেই আমরা জানি। সুতরাং ওই চেহারাটা লিখে নিই প্রথমে--

> The least square line of v on u is
>
> $$v - \bar{v} = \frac{\text{cov}(u,v)}{\text{var}(u)}(u - \bar{u}). \tag{1}$$

তার মানে বার করতে হবে $\bar{u}, \bar{v}, \text{var}(u)$ আর $\text{cov}(u,v)$. এদেরকে এবার x আর y দিয়ে লিখে নিই।

$\therefore u = x + y$and $v = x - y$,
$\therefore \bar{u} = \bar{x} + \bar{y}$ and $\bar{v} = \bar{x} - \bar{y}$.

আমরা আগে শিখেছিলাম যে, covariance জিনিসটা ঠিক গুণ করার মত আচরণ করে। এবার সেটা কাজে লাগবে--

Also

$$\begin{aligned} \text{var}(u) &= \text{var}(x+y) = \text{var}(x) + \text{var}(y) + 2\text{cov}(x,y), \\ \text{cov}(u,v) &= \text{cov}(x+y, x-y) = \cdots = \text{var}(x) - \text{var}(y). \end{aligned}$$

এবার সব কিছুই আমরা x, y দিয়ে প্রকাশ করে নিয়েছি। এখন $\bar{x}, \bar{y}, \text{var}(x), \text{var}(y)$ আর $\text{cov}(x,y)$ বার করতে পারলেই হয়।

The least square line of y on x is $y = -\frac{1}{3}x$, which has slope

$$\frac{\text{cov}(x,y)}{\text{var}(x)} = -\frac{1}{3}.$$

The least square line of x on y is $x = -\frac{2}{3}y$, which has slope

$$\frac{\text{cov}(x,y)}{\text{var}(y)} = -\frac{2}{3}.$$

লক্ষ কর যে, slope দুটো বার করার সময়ে x আর y-এর মধ্যে কখন কাকে বাঁদিকে রাখছি। যখন y on x নিয়ে কাজ করছি, তখন y থাকছে বাঁদিকে। আবার যখন x on y নিয়ে কাজ হচ্ছে, তখন বাঁদিকে থাকছে x.

The two lines intersect at $(\bar{x}, \bar{y}) = (0, 0)$. So $\bar{x} = 0$ and $\bar{y} = 0$.
It is given that $\text{var}(x) = 1^2 = 1$.
So $\text{cov}(x,y) = -\frac{1}{3} \times 1 = -\frac{1}{3}$.
Hence $\text{var}(y) = -\frac{2}{3}/\text{cov}(x,y) = 2$.

ব্যস্, এবার আবার u, v-তে ফিরে যাব--

So

$$\begin{aligned} \bar{u} &= 0 + 0, \\ \bar{v} &= 0 - 0, \\ \text{var}(u) &= 1 + 2 + 2 \times \left(-\tfrac{1}{3}\right) = \tfrac{7}{3}, \\ \text{cov}(u,v) &= 1 - 2 = -1. \end{aligned}$$

Thus, equation $(\mathbf{1})$ is

$$(v - 0) = \frac{-1}{7/3}(u - 0),$$

or

$$v = -\frac{3u}{7}.$$

■

DAY 22
Bivariate distributions (part 1)

22.1 Joint PDF এবং PMF

যদি X একটা random variable হয়, তবে তার distribution প্রকাশ করার বিভিন্ন পদ্ধতি আমরা দ্বিতীয় অধ্যায়ে শিখেছি। তাদের মধ্যে একটা ছিল PDF. এটা এমন একটা function (ধরো $f(x)$) যাতে $\{X \in A\}$ জাতীয় যেকোনো event-এরই probability বার করা যেত এইভাবে

$$P(X \in A) = \int_A f(x)dx, \quad A \subseteq \mathbb{R}.$$

এখানে $\int_A$ মানে integration-টা A-র উপরে করা হচ্ছে, যেমন $A = (0,1)$ হলে $\int_A$ মানে $\int_0^1$, এইরকম।
যদি X, Y দুটো random variable হয়, তাহলেও অনেক সময়ে একইরকম একটা জিনিস করা যেতে পারে। একে বলে joint (bivariate) probability density function. সংজ্ঞাটা আগে লিখে নিই। তারপর জিনিসটা বোঝাচ্ছি।

Definition: Joint (bivariate) PDF

Let X, Y be two jointly distributed random variables. We say that (X, Y) has joint (bivariate) PDF $f(x, y)$ if $f : \mathbb{R}^2 \to [0, \infty)$ is a function such that for any event of the form $\{(X, Y) \in A\}$ (where $A \subseteq \mathbb{R}^2$) we have

$$P((X, Y) \in A) = \iint_A f(x, y)dxdy.$$

এই সংজ্ঞায় বেশ কয়েকটা নতুন জিনিস আছে, সেগুলো এবার একে একে আলোচনা করব। তার আগে বলে নিই যে, (X, Y)-এর joint (bivariate) PDF-কে অনেক সময়ে খালি joint PDF-ও বলে (যেহেতু X আর Y দুজন আছে, তাই না বললেও "bivariate" বোঝা যাচ্ছে)। অনেক সময়ে "joint" শব্দটাও লোকে চেপে যায়।
এবার নতুন জিনিসগুলো আলোচনা করা যাক--

- সংজ্ঞাটার গোড়াতেই বলেছে "Let X, Y be two jointly distributed random variables."

  এর মানে হল X আর Y দুজনেই একই random experiment-এর ভিত্তিতে তৈরী। যেমন ধরো, একটা ক্যাসিনোতে দুরকম খেলা আছে--এক, একটা কয়েন টস করা হয়, head পড়লে তুমি একশ টাকা পাবে, tail পড়লে তুমি একশ টাকা দেবে। আর দুই, একটা ছক্কা চালা হবে, জোড় পড়লে তুমি পাঁচশ টাকা পাবে, বিজোড় পড়লে তুমি পাঁচশ টাকা দেবে। ক্যাসিনোর নিয়ম হল প্রতি সোম, বুধ, শুক্রতে

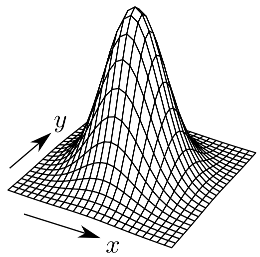


Fig 20

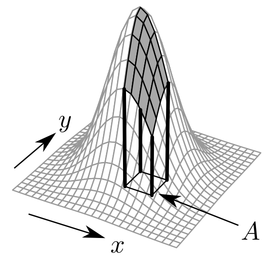


Fig 21

প্রথম খেলাটা হয়, আর সপ্তাহের বাকি দিনগুলোতে দ্বিতীয়টা। যদি প্রথম খেলার outcome-কে X আর দ্বিতীয়টার outcome-কে Y বলি, তবে তুমি বার করতে পারবে $P(X+Y=600)$ কত? না, এই probability-টার কোনো মানেই হয় না, কারণ X আর Y আলাদা দিনের জিনিস, ওদেরকে এক সঙ্গে পাওয়াই যাবে না! কিন্তু যদি খেলাটা এরকম হত--একটা কয়েন টস করা হবে এবং একই সঙ্গে একটা ছক্কাও চালা হবে, তারপর টস থেকে X এবং ছক্কা থেকে Y ঠিক করা হবে, তবে কিন্তু $X+Y$ বার করতে কোনো অসুবিধা নেই। কারণ এখানে X, Y দুজনেই একই random experiment-এর ভিত্তিতে তৈরী। এই random experiment-টা হল "একই সঙ্গে কয়েন টস করা এবং ছক্কা চালা"।

- সংজ্ঞাতে একটা function-এর কথা বলা আছে $f(x,y)$, যেখানে $f:\mathbb{R}^2 \to [0,\infty)$. এরকম function-রা দেখতে কীরকম?

  সহজ উত্তর হল--$f(x,y)$ হল একটা ফর্মুলা, যার মধ্যে x, y দুইই আছে, যেমন $(x+y)^2$ বা ye^x, এইরকম। অনেক সময়ে এদের গ্রাফটা কল্পনা করতে পারলে সুবিধা হবে। গ্রাফগুলো দেখতে কতকটা তাঁবুর মত হবে। একটা উদাহরণ দেখিয়েছি Fig 20-এ। এখানে $f(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$.

- সংজ্ঞায় লিখেছে $\iint_A f(x,y)dxdy$. সেটার মানে কী?

  এটা বোঝার জন্য $f(x,y)$-এর গ্রাফটাকে তাঁবুর মত কল্পনা কর। তাঁবুর ঢেউখেলানো ছাদটা হল গ্রাফটা, মেঝেটা হল $\mathbb{R}^2$. এখানে $A \subseteq \mathbb{R}^2$, অর্থাৎ A হল মেঝের উপর খানিকটা জায়গা। মনে মনে সেই জায়গাটার সীমানাটা চক দিয়ে এঁকে নাও। একটা উদাহরণ দেখিয়েছি Fig 21-এ। এখানে A একটা rectangle, যেটা মেঝের উপর এঁকে নিয়েছি। এবার A-র উপরে তাঁবুর ছাদ পর্যন্ত মোট যে জায়গাটা, তার volume-ই হল $\iint_A f(x,y)dxdy$. এটা কী করে বার করতে হয়, সেই প্রসঙ্গে পরে আসছি। আপাততঃ খালি মনে রাখো জিনিসটা কী।

যদি কেউ একটা function দেয় $f(x,y)$, এবং জানতে চায় সেটা একটা joint PDF হতে পারে কিনা, তবে নীচের theorem-টা দিয়ে সেই উত্তরটা দেওয়া যাবে।

THEOREM

A function $f:\mathbb{R}^2 \to \mathbb{R}$ is a bivariate PDF of some jointly distributed random variables X, Y if and only if it satisfies both the following conditions:

1. $\forall x, y \in \mathbb{R} \quad f(x,y) \geq 0$,
2. $\iint_{\mathbb{R}^2} f(x,y)dxdy = 1$.

আমরা যখন দ্বিতীয় অধ্যায়ে univariate ক্ষেত্রে PDF-এর কথা আলোচনা করেছিলাম, তখন সেই সঙ্গে PMF-এর কথাও এসেছিল। Bivariate ক্ষেত্রেও একই কাজ করা যায়--

DEFINITION: Joint (bivariate) PMF

Let X, Y be jointly distributed random variables that can only take countably many values $(x_1, y_1), (x_2, y_2), \ldots$ Then their joint (bivariate) PMF is the function $f : \mathbb{R}^2 \to [0, 1]$ defined as

$$f(x, y) = P(X = x,\ Y = y) \quad (x, y) \in \mathbb{R}^2.$$

22.2 Joint PDF বা PMF থেকে বিভিন্ন তথ্য বার করা

Joint PDF বা PMF বলা থাকলে, তা থেকে তুমি distribution-এর নাড়িনক্ষত্র সব বার করে ফেলতে পারো। এবার তার কিছু নমুনা দেখব।

22.2.1 Marginal PDF/PMF

যদি তোমাকে বলি (X, Y)-এর joint PDF হল $f(x, y)$, তবে তুমি তা থেকে অনায়াসেই X-এর pdf পেয়ে যাবে, এইভাবে--

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy.$$

একে বলব X-এর **marginal pdf.** এখানে "marginal" শব্দটা দেখে ঘাবড়ে যেও না যেন। ওটা কোনো নতুন ধরণের PDF নয়। ওতে খালি বোঝানো হচ্ছে যে $f(x, y)$-কে integrate করে ওটা বেরিয়েছে। একইভাবে Y-এর marginal PDF-টা হবে

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx,$$

যাকে বলব Y-এর marginal PDF.

Joint PDF-কে integrate করে যেমন marginal PDF বেরোলো, সেইরকম joint PMF-কে sum করে বেরোয় marginal PMF. ধরো (X, Y)-এর joint PMF হল $f(x, y)$, আর (X, Y) এই countable-সংখ্যক value-গুলো নিতে পারে, $(x_1, y_1), (x_2, y_2), \ldots$ তাহলে X-এর marginal PMF হবে

$$f_X(x) = \sum_j f(x, y_j).$$

একইভাবে Y-এর marginal PMF হবে

$$f_Y(y) = \sum_i f(x_i, y).$$

Example 12: The joint probability density function of the random variables X and Y is

$$f(x, y) = \begin{cases} K(1 - x - y) & \text{if } x \geq 0, \quad y \geq 0, \quad x + y \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Find (i) the value of K. (ii) the marginal probability density function of X. **(2009.2e)**

SOLUTION: $f(x, y)$-এর গ্রাফটা দেখিয়েছি Fig 22-এ। এখানে তাঁবুর চালটা একটা হেলানো plane. লক্ষ কর যে, তাঁবুর উচ্চতা হল K. প্রথমে অংকে আমাদের কাজ K-টা এমনভাবে বার করা, যাতে তাঁবুর নীচে মোট volume হয় 1.

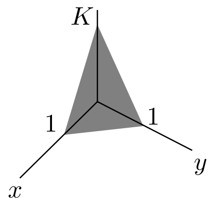


Fig 22

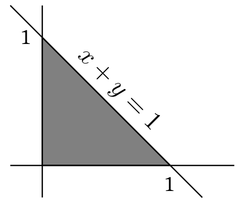


Fig 23

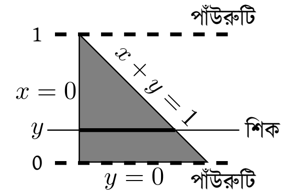


Fig 24

(i) We need $\iint_{\mathbb{R}^2} f(x,y)dxdy = 1.$
$\therefore \iint_S f(x,y)dxdy = 1,$ where $S = \{(x,y) \ : \ x \geq 0, \quad y \geq 0, \quad x+y \leq 1\}.$

এর কারণ হল S-এর বাইরে তো $f(x,y)$-টা সর্বত্রই শূন্য! S-টাই হল তাঁবুটার মেঝে। এবার আমাদের $\iint_S f(x,y)dxdy$ বার করতে হবে। এর জন্য S-এর চেহারাটার সম্বন্ধে ধারণা থাকলে সুবিধা হবে। এটা একটা ত্রিভুজ, যার ছবি রয়েছে Fig 23-এ। এবার যে কাজটা করব তাকে আমরা মজা করে নাম দিয়েছি "স্যাণ্ডউইচ-শিককাবাব কায়দা"। এর দুটো ধাপ-- স্যাণ্ডউইচ আর শিককাবাব। এর জন্য S-কে মনে করো একটা মাংসের টুকরো। Fig 24-এর দিকে চোখ রাখো। প্রথম ধাপে আমরা এটা দিয়ে স্যাণ্ডউইচ বানাব। তার জন্য দুটো horizontal লাইন নেব (এরা হল স্যাণ্ডউইচের পাঁউরুটি), এবং এদের দিয়ে S-কে উপরনীচ থেকে সাঁড়াশির মত চেপে ধরব। এখানে নীচের লাইনটা গেছে $y=0$ দিয়ে, আর উপরেরটা $y=1$ দিয়ে। তার মানে এখানে y যেতে পারে 0 থেকে 1 পর্যন্ত। সুতরাং ব্যাপারটা হবে এরকম দেখতে--

$$\iint_S f(x,y)dxdy = \int_0^1 \text{(কিছু একটা)}\, dy.$$

ওই (কিছু একটা)-টা কী হবে সেটা বোঝার জন্য পরবর্তী ধাপ, মানে শিককাবাব। শিককাবাব কী করে বানায় দেখেছো তো? একটা লোহার শিক নিয়ে তাতে মাংসের খণ্ড গেঁথে আগুনে ঝলসায়। আমাদের শিকটা হবে একটা horizontal লাইন। স্যাণ্ডউইচের দুই সীমানার মাঝামাঝি যেকোনো একটা y নিয়ে সেখান দিয়ে এই শিকটা ঢুকিয়ে দাও। যেহেতু লাইনটা horizontal, তাই এই লাইন বরাবর y স্থির আছে, খালি x পরিবর্তিত হচ্ছে। শিকটা কোনো একটা পাশ দিয়ে ঢুকে বিপরীত পাশ দিয়ে বেরিয়ে গেছে। এই দুই পাশের boundary-র equation থেকে জানা যাবে যে, এই y-এর জন্য x কোথা থেকে কতদূর যেতে পারে। Fig 24 দেখলেই বুঝবে যে, যেখান দিয়ে ঢুকেছে সেই boundary-র equation হল $x=0$. আর যেখান দিয়ে বেরিয়েছে তার equation হল $x+y=1$, বা $x=1-y$. সুতরাং x যাচ্ছে 0 থেকে $1-y$ অব্দি। অতএব ওই (কিছু একটা)-টা হবে

$$\int_0^{1-y} f(x,y)dx.$$

তার মানে সব মিলিয়ে হল

$$\iint_S f(x,y)dxdy = \int_0^1 \text{(কিছু একটা)}\, dy = \int_0^1 \left[\int_0^{1-y} f(x,y)dx\right] dy.$$

∎

এই integral-টাকে 1 হতে হবে। সেখান থেকে আমরা K বার করব।

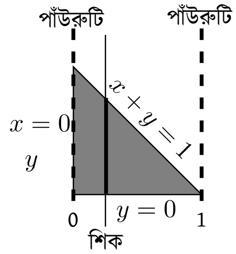


Fig 25

Thus

$$\int_0^1 \left[\int_0^{1-y} f(x,y)dx\right] dy = 1$$

or, $K\int_0^1 \left[\int_0^{1-y} (1-x-y)dx\right] dy = 1$

or, $K\int_0^1 \left[(1-y)x - \frac{x^2}{2}\Big|_{x=0}^{1-y}\right] dy = 1$

এই যে integration-টা করলাম, সেটা কিন্তু x-এর সাপেক্ষে করলাম। এখানে y স্রেফ একটা constant-এর মত আচরণ করছে।

or, $\int_0^1 \frac{1}{2}(1-y)^2 dy = 1$

$= \cdots$

or, $\frac{K}{6} = 1.$

So $K = 6.$

এবার X-এর marginal PDF বার করা তো একটা integration-এর ব্যাপার--

(ii) The marginal PDF of X is $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$

এখানে আবার স্যাণ্ডউইচ শিককাবাবের গল্প আসবে। এখানে integration-টা হচ্ছে y-এর সাপেক্ষে, তাই স্যাণ্ডউইচ আর শিককাবাব দুটোই তৈরী হবে vertical লাইন দিয়ে। Fig 25 দ্যাখো। স্যাণ্ডউইচের পাঁউরুটি দুটো আছে $x=0$ আর $x=1$-এ। সুতরাং এদের বাইরে S-এর কোনো অংশই নেই। তাই--

Clearly, $\int_{-\infty}^{\infty} f(x,y)dy = 0$ if $x \notin (0,1)$.

যদি $x \in (0,1)$ হয়, তবে শিক ঢোকাব x দিয়ে। সেটা $y=0$-র পাশ দিয়ে ঢুকে $x+y=1$-এর পাশ ফুঁড়ে বেরিয়ে আসবে। তার মানে এই x-এর জন্য y যেতে পারে 0 থেকে $1-x$ অব্ধি--

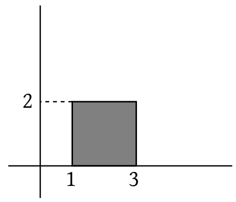


Fig 26

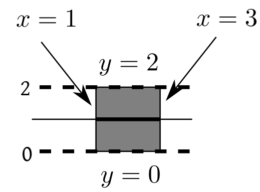


Fig 27

> If $x \in (0, 1)$, then
> $$\int_{-\infty}^{\infty} f(x, y)dy = \int_0^{1-x} 6(1 - x - y)dy = \cdots = 3(1 - x)^2.$$

এই integration-টার সময়ে x-টা কিন্তু constant-এর মত আচরণ করছে।

> Thus the required density is
> $$f_X(x) = \begin{cases} 3(1-x)^2 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

■

এই অংকটা ভালো করে পড়ে স্যাণ্ডউইচ-শিককাবাবের কায়দাটা বুঝে নাও। এটা ফের কাজে লাগবে নীচের অংকটায়।

Example 13: If the joint probability density function of the random variables X, Y is

$$f(x, y) = \begin{cases} k(3x + y) & \text{if } 1 \le x \le 3, \quad 0 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$$

find the value of k. Find (i) $P(X + Y < 2)$, (ii) the marginal distribution of X and Y.[1+2+2] **(2014.3b)**

SOLUTION:

> We know that $\iint_{\mathbb{R}^2} f(x, y)dxdy = 1$.
> So $\iint_S k(3x + y)dxdy = 1$, where
> $$S = \{(x, y) \ : \ 1 \le x \le 3, \quad 0 \le y \le 2\}.$$

এখানে S হল একটা square, যেটাকে দেখিয়েছি Fig 26-এ। এর উপর স্যাণ্ডউইচ-শিককাবাব করলে কী হয়, সেটা রয়েছে Fig 27-এ। এই ছবিটা মাথায় রেখে নীচের ধাপগুলো পড়।

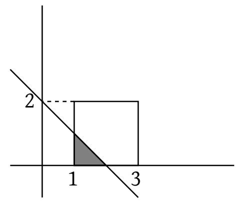


Fig 28

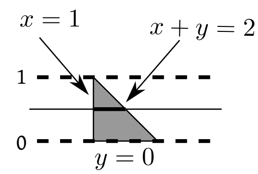


Fig 29

Now

$$\iint_S k(3x+y)dxdy = k\int_0^2 \left[\int_1^3 (3x+y)dx\right] dy$$

ভিতরের integral-টা কর y-কে constant ধরে নিয়ে।

$$= \cdots$$

$$= k\int_1^3 2y + 12\,dy = \cdots = 28k.$$

Thus $28k = 1$, and so $k = \frac{1}{28}$.

এইবার পরের অংশ--

Let $A = \{(x,y) \in S \ : \ x+y < 2\}$.

আমাদের $P((X,Y) \in A)$ বার করতে বলেছে। এই A-এর ছবি এঁকেছি Fig 28-এ। স্যাণ্ডউইচ-শিককাবাব করার জন্য আরেকটু বড় করে A-কে দেখিয়েছি Fig 29-এ।

Then

$$P(X+Y<2) = P((X,Y)\in A) = \iint_A f(x,y)dxdy$$

$$= k\int_0^1 \left[\int_1^{2-y} (3x+y)dx\right] dy$$

$$= \cdots = \frac{13}{168}.$$

■

22.2.2 Independence পরীক্ষা করা

আমরা তৃতীয় অধ্যায়ে দুটো event-এর independence-এর সংজ্ঞা শিখেছিলাম। দুটো random variable-কে কখন independent বলে সে বিষয়েও আমাদের একটা ধারণা হয়েছে। এবার দেখব joint PDF বা PMF ব্যবহার করে কীকরে independence পরীক্ষা করতে হয়। প্রথমে দুটো random variable-এর independence-এর সংজ্ঞাটা গুছিয়ে লিখে নিই--

Definition: Independence of random variables

Two jointly distributed random variables X, Y are called <u>independent</u> if for all events of the form $\{X \in A\}$ and $\{Y \in B\}$ (where $A, B \subseteq \mathbb{R}$) we have

$$P(X \in A \text{ and } Y \in B) = P(X \in A)P(Y \in B).$$

অবশ্য এই সংজ্ঞা ব্যবহার করে দুটো random variable-এর independence পরীক্ষা করা কঠিন, কারণ বিশ্বের যাবতীয় $\{X \in A\}$আর $\{Y \in B\}$ জাতীয় event নিয়ে কাজ করা কঠিন। কিন্তু যদি (X, Y)-এর joint pdf বা PMF থাকে তবে কাজটা অনেকটাই সোজা। কারণ সেখানে এই theorem-টা আছে--

Theorem

Let X, Y have joint PDF/PMF $f(x, y)$ and marginal PDFs/PMFs $f_X(x)$and $f_Y(y)$. If

$$\forall x, y \quad f(x, y) = f_X(x)f_Y(y),$$

then X, Y are independent.

PMF-এর বেলায় এর উল্টোদিকটাও জোর দিয়ে বলা যায়। দুঃখের কথা, PDF-এর বেলায় সেটা করা যায় না। কিন্তু যদি PDF-গুলো কোথাও continuous হয়, তবে উল্টো দিকটাও বলা যায়, এইভাবে--

Theorem

Let X, Y have joint PDF $f(x, y)$ and marginal PDFs $f_X(x)$ and $f_Y(y)$. Let $f_X(x)$ be continuous at $x = a$. Let $f_Y(y)$ be continuous at $y = b$. Let $f(x, y)$ be continuous at $(x, y) = (a, b)$. If X, Y are independent, then

$$f(a, b) = f_X(a)f_Y(b).$$

এটা আমরা এখানে প্রমাণ করব না।
একটা প্রয়োগ রয়েছে নীচের অংকটায়--

Exercise 2: The p.d.f. of a two dimensional r.v. is

$$f(x, y) = \frac{1}{8}(x + y); \quad 0 \le x \text{ and } y \le 2.$$

Are X and Y independent?[4] **(2013.3a)**

Hint:

এখানে r.v. মানে হল random variable. অংকটা পুরোটা করে দেব না, খালি মূল ধাপগুলো বলে দিচ্ছি। প্রথমে X-এর marginal PDF বার কর। দেখবে উত্তর আসছে--

$$f_X(x) = \begin{cases} \frac{1}{4}(x+1) & \text{if } x \in [0,2] \\ 0 & \text{otherwise} \end{cases}$$

একবার $f_X(x)$ বেরিয়ে গেলে, খালি x-গুলোর জায়গায় y বসালেই $f_Y(y)$-ও পেয়ে যাবে, কারণ এখানে $f(x,y)$ হল x, y-এর সাপেক্ষে symmetric (অর্থাৎ $f(x,y) = f(y,x)$)।

By symmetry, the marginal PDF of Y is

$$f_Y(y) = \begin{cases} \frac{1}{4}(y+1) & \text{if } y \in [0,2] \\ 0 & \text{otherwise} \end{cases}$$

বুঝতেই পারছ যে, $f(x,y)$ কোনো ভাবেই $f_X(x)f_Y(y)$ নয়। গুছিয়ে লেখার জন্য এমন a, b বার কর যাতে $f(a,b) \neq f_X(a)f_Y(b)$ হয়, এবং a, b-তে PDF-গুলো সবাই continuous-ও হয়। $a = 1/2$ আর $b = 1/2$ নিলেই হবে। এরকম আরও বহু a, b সম্ভব।

Clearly, $f(x,y), f_X(x), f_Y(y)$ are continuous at $(x,y) = \left(\frac{1}{2}, \frac{1}{2}\right), x = \frac{1}{2}, y = \frac{1}{2}$, respectively. But $f\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{8} \neq f_X\left(\frac{1}{2}\right) f_Y\left(\frac{1}{2}\right) = \frac{9}{64}$.
$\therefore X, Y$ are not independent.

■

DAY 23
Bivariate distributions (part 2)

23.1 Joint PDF/PMF থেকে আরও তথ্য বার করা

23.1.1 Conditional PDF/PMF

যদি একটা unbiased কয়েন নিয়ে 10 বার টস করি, তবে মোট head-এর সংখ্যার distribution যে Binomial$\left(10, \frac{1}{2}\right)$ হবে সে তো জানোই।

এবার এরই একটু কঠিন সংস্করণ দেখি। ধরো একটা ছক্কা চাললাম। যত সংখ্যা আসবে, তত গুণ দশবার একটা কয়েন টস করা হবে। যেমন ছক্কায় যদি 2 পড়ে, তবে $2 \times 10 = 20$ বার টস করব। এবার ধরো তোমাকে বলে দিলাম ছক্কায় 3 পড়েছে। এটুকু মাত্র শুনে বলতে হবে মোট head-এর সংখ্যার distribution কী হবে। যেহেতু ছক্কায় 3 পড়েছে, সুতরাং টস করা হবে $3 \times 10 = 30$ বার। সুতরাং বুঝতেই পারছ যে, উত্তর হবে Binomial$\left(30, \frac{1}{2}\right)$. এখানে কাজটা কী করলাম দেখি। ধরো ছক্কার outcome-টাকে বললাম X আর মোট head-এর সংখ্যাকে বললাম Y. তাহলে $X = 3$ বলে দিয়ে Y-এর distribution বার করতে বলা হয়েছিল। এই distribution-টাকে আমরা বলি **conditional distribution of Y given X.** লেখার সময়ে লিখি $Y|X = 3 \sim$ Binomial$\left(30, \frac{1}{2}\right)$. যদি X-এর অন্য কোনো value (ধরো x) দেওয়া থাকত তবে উত্তরটা হত Binomial$\left(10x, \frac{1}{2}\right)$. তাই আমরা লিখব

$$\forall x \in \{0, 1, ..., 6\} \quad Y|X = x \sim \text{Binomial}\left(10x, \frac{1}{2}\right).$$

কোনো distribution-এর যেমন PDF বা PMF থাকতে পারে, conditional distribution-এরও তাই। এর PDFবা PMF-কে বলে conditional PDF বা conditional PMF.

Definition: Conditional PDF/PMF

Let X, Y be jointly distributed random variables with joint PDF (or PMF) $f(x, y)$. Let the marginal PDFs (or PMFs) of X and Y be $f_X(x)$ and $f_Y(y)$, respectively. Let x_0, y_0 be two numbers such that $f_X(x_0) > 0$ and $f_Y(y_0) > 0$. Then the conditional PDF (PMF) of Y given $X = x_0$ is defined as

$$f_{Y|X}(y|x_0) = \frac{f(x_0, y)}{f_X(x_0)}.$$

Similarly, the conditional PDF (PMF) of X given $Y = y_0$ is defined as

$$f_{X|Y}(x|y_0) = \frac{f(x, y_0)}{f_Y(y_0)}.$$

এমনি PDF/PMF-এর যা যা গুণ, conditional PDF/PMF-এরও তাই। PMF-কে sum করে যেমন probability বার করা যায়, তেমনি conditional PMF-কে sum করে পাওয়া যায় conditional probability. ওদিকে, PDF-কে integrate করে যেমন probability বার করা যায়, তেমনি conditional PDF-কে integrate করলেও পাওয়া যায় conditional probability. এমনি PDF/PMF থেকে যেভাবে expectation, variance ইত্যাদি বেরোয়, সেই একইভাবে conditional PDF/PMF থেকে বেরোয় conditional expectation আর conditional variance.
এবার আমরা এইভাবে conditional probability বার করার একটা উদাহরণ দেখব। তবে conditional PDF-কে integrate করে যে conditional probability বেরোয়, তার সঙ্গে আমরা তৃতীয় অধ্যায়ে যে conditional probability-র কথা শিখেছিলাম, তার একটা গুরুত্বপূর্ণ পার্থক্য আছে। আগে সেটা বুঝে নিই। যদি $\int_1^2 f_X(x)dx$ বার করি, তবে সেটা কোন probability? উত্তর হল $P(X \in [1, 2])$. তেমনি যদি $\int_1^2 f_{X|Y}(x|3)dx$ বার করি সেটা হবে $P(X \in [1, 2]|Y = 3)$. এটা দেখতে ঠিক $P(A|B)$-এর মতন, যেখানে $A = \{X \in [1, 2]\}$ আর $B = \{Y = 3\}$. তৃতীয় অধ্যায়ে $P(A|B)$ মানে শিখেছিলাম--

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

এবং এর জন্য $P(B) > 0$ হওয়া দরকার ছিল। সেই সংজ্ঞা কিন্তু এখানে লাগানো যাবে না, কারণ এখানে $P(B) = 0$ আসবে (যেহেতু Y-এর density আছে, তাই singleton set-এর probability শূন্য, $P(Y = 3) = \int_3^3 f_Y(y)dy = 0.$)। তবে উপায়? আসলে এখানে "$Y = 3$" মানে হল "Y-কে মেপে পাওয়া গেছে 3." যেহেতু যেকোনো পরিমাপেই কিছুটা approximation থাকে, তাই আসলে $Y = 3$ মানে $Y \in [3 - \delta, 3 + \delta]$, যেখানে $\delta > 0$ হল পরিমাপের সূক্ষ্মতার উপর নির্ভরশীল, যেমন যদি কোনো জিনিসের দৈর্ঘ্য মেপে 4.7 cm পাওয়া যায়, যেখানে স্কেলে মিলিমিটার অব্ধি দাগ কাটা আছে, তবে বুঝতে হবে যে, প্রকৃত দৈর্ঘ্য হল 4.65 cm থেকে 4.75 cm-র মধ্যে কিছু একটা। এখানে $\delta > 0$ যতই ছোটো হবে, ততই আমরা সঠিক উত্তরের কাছাকাছি এগিয়ে যাব। তাই $P(X \in [1, 2]|Y = 3)$-এর সংজ্ঞা নেওয়া হয় এইভাবে--

$$P(X \in [1, 2]|Y = 3) = \lim_{\delta \to 0+} P\big(X \in [1, 2]\big|Y \in [3 - \delta, 3 + \delta]\big).$$

সংজ্ঞাটা দেখতে যতই খটমট লাগুক, বার করার কায়দাটা সহজই--

$$P(X \in [1, 2]|Y = 3) = \int_1^2 f_{X|Y}(x|3)dx.$$

একটা উদাহরণ দেখা যাক।

Example 14: The joint probability density function of two variables X, Y is given by

$$f(x,y) = \begin{cases} \frac{(6-x-y)}{8} & \text{if } 0 < x < 2; \quad 2 < y < 4 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the following probabilities: (i) $P(X < 1,\ Y < 3)$, (ii) $P(X < 1|Y = 3)$.[2+3] **(2011.3b)**
SOLUTION: প্রথম অংশটা খুবই সহজ। তোমার করার জন্য রেখে দিলাম। বস্তুতঃ এখানে খালি এই double integral-টা বার করতে হবে--

$$\int_2^3 \left[\int_0^1 \frac{(6-x-y)}{8} dx\right] dy = \cdots = \frac{3}{8}.$$

দ্বিতীয় অংশে একটা conditional probability বার করতে বলেছে $P(X < 1|Y = 3)$. প্রথমে conditional PDF বার করতে হবে, $f_{X|Y}(x|3)$. তারপর সেটাকে x-এর উপর integrate করতে হবে $-\infty$ থেকে 1 পর্যন্ত, ব্যস্‌!

The marginal PDF of Y at $y = 3$ is

$$f_Y(3) = \int_{-\infty}^{\infty} f(x,3)dx = \int_0^2 \frac{(6-x-3)}{8} dx = \frac{1}{2}.$$

So the conditional PDF of X given $Y = 3$ is

$$f_{X|Y}(x|3) = \frac{f(x,3)}{f_Y(3)} = 2f(x,y) = \begin{cases} \frac{(3-x)}{4} & \text{if } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

So

$$P(X < 1|Y = 3) = \int_{-\infty}^1 f_{X|Y}(x|3)dx = \int_0^1 \frac{3-x}{4} dx = \frac{5}{8}.$$

■

23.2 Uniform distribution

আমরা গতকাল শিখেছি কীভাবে bivariate PDF-দের স্যাণ্ডউইচ-শিককাবাব করে integrate করতে হয়। কায়দাটা খুব কঠিন নয়, কিন্তু একটু লম্বা। একটা বিশেষ ক্ষেত্রে কাজটা কিন্তু অনেক সহজে করা যায়--যখন PDF-টা constant হয়, মানে কোনো একটা set থাকে S, যাতে PDF-টার চেহারা হয় এইরকম--

$$f(x,y) = \begin{cases} k & \text{if } (x,y) \in S \\ 0 & \text{otherwise} \end{cases}$$

যেখানে $k > 0$ একটা constant আর $S \subseteq \mathbb{R}^2$. সহজ ভাষায় এর মানে হল (X,Y) খালি S-এর মধ্যেই value নিতে পারে, এবং কোনো value-র প্রতিই বেশী বা কম পক্ষপাত নেই। এখানে আমরা বলি যে, "(X,Y) is uniformly distributed over S." এইরকম অংক আমরা আগেই দেখেছিলাম প্রথম অধ্যায়ে, খালি সেখানে S ছিল finite. সেখানকার একটা উদাহরণ ছিল এইরকম--একটা ছক্কা চালা হল (যার সব দিক সমান ভারী), তাহলে even (জোড়) সংখ্যা আসার probability কত? সেখানে $S = \{1,2,3,4,5,6\}$ আর event-টা ছিল $A = \{2,4,6\}$, যেখানে A হল যাবতীয় even সংখ্যার set. উত্তর বার করার কায়দাটা ছিল--

$$P(A) = \frac{|A|}{|S|}.$$

সেখানে $|S|$ আর $|A|$ বার করা যাচ্ছিল কারণ S একটা finite set ছিল। কিন্তু যদি S একটা infinite set-ও হয়, তাও একইরকমের একটা জিনিস এখানে করা যাবে, যদি distribution-টা S-এর উপরে uniform হয়। সেক্ষেত্রে--

$$P(A) = \frac{A \cap S\text{-এর area}}{S\text{-এর area}}.$$

এতে সুবিধা হল কোনো volume বার করতে হয় না, খালি area বার করলেই হয়। যেহেতু volume লাগছে না, তাই double integral-ও কষতে হয় না। সাধারণতঃ area বার করা অনেক বেশী সহজ, বিশেষতঃ যদি ত্রিভুজ, square বা circle-জাতীয় পরিচিত জিনিসের area বার করা হয়। এবার এরকম দুটো উদাহরণ দেখি।

Example 15: Two numbers are independently chosen at random between 0 and 1. Show that the probability that their product is less than a constant k $(0 < k < 1)$ is $k(1 - \log k)$. **(2007,2005)**

SOLUTION: সংখ্যা দুটোর নাম দেওয়া যাক Xআর Y. আমাদের প্রথম কাজ হল X,Y-এর joint PDF-টা বার করা। যেহেতু X, Y বলা আছে independent, তাই joint PDF-টা হবে ওদের PDF দুটোর গুণফল।

> Let X, Y be IID Unif$(0, 1)$.
> Then X has PDF
> $$f_X(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$
> and Y has PDF
> $$f_Y(y) = \begin{cases} 1 & \text{if } 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$
> Since X, Y are independent, their joint PDF is
> $$f(x, y) = f_X(x) f_Y(y) = \begin{cases} 1 & \text{if } 0 < x, y < 1 \\ 0 & \text{otherwise} \end{cases},$$
> Let $S = \{(x, y) \ : \ 0 < x, y < 1\}$.

লক্ষ কর যে, এই joint PDF-টা S-এর বাইরে 0, এবং S-এর ভিতরে constant. সুতরাং এখানে volume বার না করে area বার করলেই চলবে।

> Then (X, Y) is distributed uniformly over S.

Fig 30

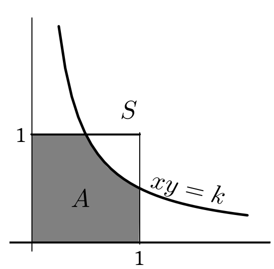
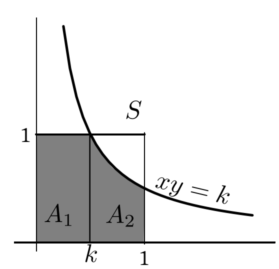


Fig 31



> Hence for any event $A \subseteq S$, we have
> $$P((X,Y)\in A) = \frac{\text{area}(A)}{\text{area}(S)}.$$

এইবার আমাদের অংকে কী event দিয়েছে দেখি--

> Let $k \in (0,1)$.
> Then we want to find $P((X,Y)\in A)$, where
> $$A = \{(x,y)\in S \ : \ xy < k\}.$$

এবার S আর A-র area বার করতে হবে। Fig 30 দ্যাখো। এখানে square-টা হল S, আর তার মধ্যে শেড করা অংশটা হল A.

> Clearly, area$(S) = 1$, because S is a square with sides of length 1.
> Also
> $$\begin{aligned}\text{area}(A) &= \text{area}(A_1) + \text{area}(A_2) \quad [\ A_1, A_2 \text{ as shown.}]\\ &\qquad \text{এখানে উত্তরের মধ্যে Fig 31-এর মত ছবি এঁকে বোঝানোটা অপরিহার্য।}\\ &= k + \int_k^1 \tfrac{k}{x}dx\\ &\qquad \text{যেহেতু } A_1 \text{ একটা rectangle, যার উচ্চতা 1 এবং প্রস্থ } k.\\ &= k - k\log k.\end{aligned}$$
> Thus, $P(A) = \frac{\text{area}(A)}{\text{area}(S)} = k(1-\log k)$, as required.

■

এবার আরেকটা একই রকম অংক।

Example 16: If p and q are independent variates each uniformly distributed over the interval $(-1,1)$, find the probability that the equation $x^2 + 2px + q = 0$ has real roots. **(2007.2c)**

SOLUTION:

> Let p,q be IID Unif$(-1,1)$.
> Then p has PDF
> $$f_p(p) = \begin{cases} \frac{1}{2} & \text{if } -1<p<1\\ 0 & \text{otherwise}\end{cases}$$
> and q has PDF
> $$f_q(q) = \begin{cases} \frac{1}{2} & \text{if } -1<q<1\\ 0 & \text{otherwise}\end{cases}$$

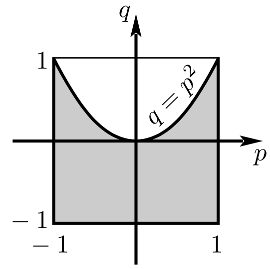


Fig 32

Since p, q are independent, their joint PDF is

$$f(p,q) = f_p(p)f_q(q) = \begin{cases} \frac{1}{4} & \text{if } -1 < p, q < 1 \\ 0 & \text{otherwise} \end{cases}.$$

Let $S = \{(p,q) \ : \ -1 < p, q < 1\}$.

Then (X, Y) is distributed uniformly over S.

সুতরাং আগের অংকের মত এখানেও volume বার না করে area বার করলেই চলবে।

Hence for any event $A \subseteq S$, we have

$$P((X,Y) \in A) = \frac{\text{area}(A)}{\text{area}(S)}.$$

এইবার আমাদের অংকের event-টা কী দেখি--

The discriminant of the given equation is $4p^2 - 4q$.

So the equation will have real roots if and only if $p^2 \geq q$.

Thus we want to find $P(A)$ where

$$A = \{(p,q) \in S \ : \ p^2 \geq q\}.$$

এবার S আর A-এর areaবার করার পালা। Fig 32-এ ওদের ছবি রয়েছে। Square-টা হল S, আর তার মধ্যে শেড করা অংশটা A.

Since S is a square with sides of length 2, so area$(S) = 4$.

Also

$$
\begin{aligned}
\text{area}(A) &= \text{area}(A_1) + \text{area}(A_2) \\
&\quad \text{where } A_1 \text{ is the part of } A \text{ below } x\text{-axis,} \\
&\quad \text{and } A_2 \text{ is the remaining part.} \\
&\quad \text{ছবি আঁকাটা খুবই দরকারী এখানে।} \\
&= 2 + \int_{-1}^{1} p^2 dp \quad [\because \text{area}(A_1) = 2] \\
&= 2 + \frac{1}{3} p^3 |_{-1}^{1} \\
&= 2 + \tfrac{2}{3} = \tfrac{8}{3}.
\end{aligned}
$$

So the required probability is

$$P(A) = \frac{\text{area}(A)}{\text{area}(S)} = \frac{8/3}{4} = \frac{2}{3}.$$

■

DAY 24
Bivariate distributions (part 3)

24.1 Expectation

আমরা univariate distribution-এর বেলায় expectation বার করা শিখেছিলাম। একইরকম কাজ করা যায় bivariate distribution-এর ক্ষেত্রেও। সংজ্ঞাগুলো একেবারেই একইরকম--

DEFINITION: Expectation (bivariate PMF version)

Let X, Y be jointly distributed random variables that can only take values in the set $S = \{(x_1, y_1), (x_2, y_2), ...\}$. Let the joint PMF be $f(x, y)$. Then for any function $h : S \to \mathbb{R}$ the expectation of $h(X, Y)$ is defined as

$$E(h(X, Y)) = \sum_i h(x, y) f(x, y),$$

provided $\sum_i |h(x, y)| f(x, y) < \infty$.

যদি PDF থাকে তবেও সংজ্ঞাটা একইরকম--

Definition: Expectation (bivariate PMF version)

Let X, Y be jointly distributed random variables with joint PDF $f(x, y)$. Then for any function $h : \mathbb{R}^2 to \mathbb{R}$ the expectation of $h(X, Y)$ is defined as

$$E(h(X,Y)) = \iint_{\mathbb{R}^2} h(x,y) f(x,y) dxdy,$$

provided $\iint_{\mathbb{R}^2} |h(x,y)| f(x,y) dxdy < \infty$.

PMF-ই থাক বা PDF-ই থাক, expectation-এর কিছু ধর্ম সব ক্ষেত্রেই একই। তাদের মধ্যে অন্যতম হল এর linearity, অর্থাৎ সহজ ভাষায় যাকে বলে "constant বেরিয়ে আসে, এবং যোগের উপর ভেঙে যায়।" অংকের ভাষায়--

Theorem

If X, Y are jointly distributed and $h_1(x, y), h_2(x, y)$ are two functions such that $E(h_1(X, Y))$and $E(h_2(X, Y))$ exist, then for any two numbers $a_1, a_2 \in \mathbb{R}$

$$E\big(a_1 h_1(X,Y) + a_2 h_2(X,Y)\big) = a_1 E\big(h_1(X,Y)\big) + a_2 E\big(h_2(X,Y)\big).$$

24.2 Correlation

দুটো random variable যদি jointly distributed হয়, তবে তাদের দুজনের মধ্যে কোনো সম্পর্ক আছে কিনা সেটা নিয়ে স্বভাবতঃই আমাদের আগ্রহ থাকে। সে কথা আমরা আগেই বলেছি। এই রকম কোনো সম্পর্ক আছে কিনা সেটা বার করার একটা হাতিয়ার হল correlation. এই ধারণাটার সঙ্গে আমাদের ইতিমধ্যেই একবার মোলাকাত হয়েছে statistics-এর দৃষ্টিভঙ্গী থেকে। এবার দেখব probability-র দৃষ্টিভঙ্গী থেকে। সংজ্ঞাটা প্রায় সেরকমই।

Definition: Covariance,correlation

Let X, Y be two jointly distributed random variables. Then their covariance is defined as

$$\mathrm{cov}(X,Y) = E\big((X - E(X))(Y - E(Y))\big).$$

If this expectation does not exist, then we say that $\mathrm{cov}(X, Y)$ is undefined.
The correlation of X, Y is defined as

$$\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}}.$$

If at least one of $\mathrm{var}(X), \mathrm{var}(Y)$ or $\mathrm{cov}(X, Y)$ does not exist or if $\mathrm{var}(X)$ or $\mathrm{var}(Y)$ is zero, then we say that $\mathrm{corr}(X, Y)$ is undefined.

চট্ করে বুঝে নেওয়া যাক যে, কদিন আগে আমরা statistics-এর দৃষ্টিভঙ্গী থেকে (sample) covariance আর (sample) correlation-এর যে সংজ্ঞা দিয়েছিলাম তার সঙ্গে এই দুটো সংজ্ঞার সম্পর্ক কী। ধরো X, Y হল jointly distributed

দুটো random variable. যে random experiment-এর সঙ্গে এরা দুজনেই জড়িত, সেই random experiment-টা বার বার করে করা যাক। ধরো 1000 বার করলাম (independently)। তাহলে X-এর 1000-টা value আর Y-এরও 1000-টা value পাবে,--

$$(X_1, Y_1), ..., (X_{1000}, Y_{1000}).$$

এই data-র উপর statistical সংজ্ঞাটা লাগাও, তবে পাবে তাদের sample correlation–

$$\frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}.$$

যদি random experiment-টা আরও আরও বেশীবার করতে থাকো তবে এই sample correlation-এর value-ও ক্রমশঃ বদলাবে। প্রমাণ করা যায় যে, এই value-টা ক্রমশঃ একটা limit-এর দিকে এগিয়ে যাবে, এবং সেই limit-টাই হল probability-র দৃষ্টিভঙ্গী থেকে দেওয়া সংজ্ঞাটা। একইরকম কথা খাটে covariance-এর ক্ষেত্রেও। সেই অর্থে probability-র দৃষ্টিভঙ্গী থেকে দেওয়া covariance আর correlation-এর statistical image হচ্ছে sample covariance আর sample correlation. তাই sample covariance-এর বিভিন্ন গুণ এখানেও থাকবে। এরকম দুটো জিনিস রয়েছে নীচের theorem দুটোয়।

THEOREM

If X, Y are jointly distributed random variables, then

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y).$$

THEOREM

Let X, Y are jointly distributed random variables, and let $a, b, c, d \in \mathbb{R}$ be fixed numbers. Then

1. $\text{cov}(X, Y) = \text{cov}(Y, X)$
2. $\text{cov}(X, X) = \text{var}(X)$
3. $\text{cov}(aX + bY, cX + dY) = ac\text{var}(X) + (ad + bc)\text{cov}(X, Y) + bd\text{var}(Y)$.
4. $\text{cov}(aX + b, cY + d) = ac\text{cov}(X, Y)$
5. $\text{corr}(aX + b, cY + d) = \text{corr}(X, Y)$.

24.2.1 কী কী value নিতে পারে

আমরা এবার দেখব যে, $\text{corr}(X, Y)$ সর্বদা -1 থেকে 1-এর মধ্যে থাকতে বাধ্য। তার প্রমাণ রয়েছে নীচের অংকটায়।

Example 17: Let (X, Y) be two-dimensional random variable. Prove that

$$\{E(XY)\}^2 \leq E(X)^2 E(Y)^2.$$

Hence deduce that

$$-1 \leq \rho(X,Y) \leq 1,$$

where ρ is the correlation coefficient between X and Y.[3+2] **(2011,2007)**

SOLUTION: (X,Y)-কে একটা two dimensional random variable বলার মানে হল X,Y হচ্ছে jointly distributed.

এই অংকটা আসলে Cauchy-Schwartz inequality-র একটা রূপ। অবশ্য এখানে আমরা Cauchy-Schwartz inequality-র উল্লেখ না করে সেটার প্রমাণও এর মধ্যে ঢুকিয়ে দেব। এর জন্য একটা কৌশল লাগবে। ঠাণ্ডা মাথায় পড়।

We have

$$E\big(XE(Y^2) - YE(XY)\big)^2 \geq 0 \quad [\because \text{square of a real number is always} \geq 0.]$$

এইটাই হল কৌশলটা। এই অদ্ভুত জিনিসটা নিয়ে শুরু করলেই কাজটা হবে।

or, $$E\left[X^2\big(E(Y^2)\big)^2 + \big(E(XY)\big)^2 Y^2 - 2XYE(Y^2)E(XY)\right] \geq 0$$

এটা পেলাম $(a-b)^2$-এর ফর্মুলা লাগিয়ে।

or, $$E(X^2)\big(E(Y^2)\big)^2 + \big(E(XY)\big)^2 E(Y^2) - 2E(Y^2)\big(E(XY)\big)^2 \geq 0$$

by linearity of expectation

মনে রেখো যে, $E(Y^2)$ ইত্যাদিরা কিন্তু constant, তাই ওরা expectation-এর বাইরে বেরিয়ে আসবে, যেমন $E(X^2E(Y^2)) = E(X^2)E(Y^2)$.

or, $$E(X^2)\big(E(Y^2)\big)^2 \geq \big(E(XY)\big)^2 E(Y^2).$$

এবার নিশ্চয়ই দুপাশ থেকে $E(Y^2)$ কেটে নিতে ইচ্ছে করছে? সাবধান $E(Y^2)$ কিন্তু 0 হতে পারে। সেই অনুযায়ী এখানে দুটো কেস হবে।

Case 1: If $E(Y^2) \neq 0$, then we can divide both sides by $E(Y^2)$ to get

$$E(X^2)E(Y^2) \geq \big(E(XY)\big)^2, \qquad (*)$$

as required.

যদি $E(Y^2) = 0$ তবে অবশ্যই এই ভাগটা করা যাবে না। কিন্তু তাতেও অসুবিধা নেই। কারণ--

Case 2: If $E(Y^2) = 0$, then $P(Y=0) = 1$.

⟦ Because:

Then $\mathrm{var}(Y) = E(Y^2) - \big(E(Y)\big)^2 = -\big(E(Y)\big)^2 \leq 0$.
But $\mathrm{var}(Y) \geq 0$. and hence $\mathrm{var}(Y) = 0$.
So Y must be a constant, k, say. So $E(Y^2) = k^2 = 0$ (given).
So $k = 0$.

⟧

এই যুক্তিটা ভালো করে পড়ে বুঝে নাও।

> So (*) again holds as both sides are equal to 0.

এবার অংকের দ্বিতীয় অংশ--

> Now putting $X-E(X)$ in place of X, and $Y-E(Y)$ in place of Y we have
> $$[E((X-E(X))(Y-E(Y))]^2 \leq E((X-E(X))^2)E((Y-E(Y))^2),$$
> or
> $$(\mathrm{cov}(X,Y))^2 \leq \mathrm{var}(X)\mathrm{var}(Y).$$
> So
> $$\frac{(\mathrm{cov}(X,Y))^2}{\mathrm{var}(X)\mathrm{var}(Y)} \leq 1,$$
> or
> $$\big(\rho(X,Y)\big)^2 \leq 1,$$
> or
> $$-1 \leq \rho(X,Y) \leq 1,$$
> as required.

■

তাহলে আমরা দেখলাম যে $\mathrm{corr}(X,Y)$ সর্বদা $[-1,1]$-এর মধ্যে থাকে। এর মধ্যে তিনটে value বিশেষভাবে গুরুত্বপূর্ণ, এরা হল $-1, 1$ এবং 0. কখন $\mathrm{corr}(X,Y)=1$ আর $\mathrm{corr}(X,Y)=-1$ হয় তার হদিশ রয়েছে নীচের অংকটায়।

Example 18: If the random variables X and Y are connected by the linear relation $aX+bY+c=0$, prove that the correlation coefficient between X and Y is -1 if a,b have the same sign, and 1 if a,b have opposite signs. **(2009.3d)**

SOLUTION: একটা উদাহরণ নিয়ে বুঝে নাও। ধরো $2X-3Y+4=0$, অর্থাৎ $a=2, b=-3$ আর $c=4$. যেহেতু এখানে a আর b-এর বিপরীত চিহ্ন, তাই প্রশ্নটা বলছে যে, $\mathrm{corr}(X,Y)=1$ হবে। কিন্তু যদি $2X+3Y+4=0$ হত, $a=2, b=3$-এর একই চিহ্ন, তাই সেখানে $\mathrm{corr}(X,Y)=-1$ হত।

> We assume that $a,b \neq 0$.

এইটা না ধরলে ওদের sign নিয়ে আলোচনা করা যাবে না। তাছাড়া a,b দিয়ে মাঝেমাঝে ভাগও করতে হবে।

> So $Y=-\frac{c}{b}-\frac{a}{b}X=\alpha+\beta X$, say, where $\alpha=-\frac{c}{b}$ and $\beta=-\frac{a}{b}$.

এবার আমাদের $\mathrm{corr}(X,Y)$ বার করতে হবে। তার জন্য বার করা দরকার $\mathrm{var}(X), \mathrm{var}(Y)$ আর $\mathrm{cov}(X,Y)$. আমরা $\mathrm{var}(Y)$ আর $\mathrm{cov}(X,Y)$-কে $\mathrm{var}(X)$ দিয়ে প্রকাশ করব--

> Now $\mathrm{var}(Y)=\mathrm{var}(\alpha+\beta X)=\beta^2\mathrm{var}(X)$.
> Also $\mathrm{cov}(X,Y)=\mathrm{cov}(X,\alpha+\beta X)=\beta\mathrm{cov}(X,X)=\beta\mathrm{var}(X)$.

So

$$\text{corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{\beta\text{var}(X)}{\sqrt{\beta^2(\text{var}(X))^2}} = \frac{\beta\text{var}(X)}{|\beta|\text{var}(X)} = sign(\beta).$$

এইখানে মনে রেখো যে $\sqrt{\beta^2} = |\beta|$. যেমন $\beta = -5$, হলে $\sqrt{\beta^2} = \sqrt{(-5)^2} = \sqrt{25} = 5 = |\beta|$.

Now, since $\beta = -\frac{a}{b}$, so $sign(\beta) = \begin{cases} 1 & \text{if } a, b \text{ have opposite signs} \\ -1 & \text{if } a, b \text{ have same sign} \end{cases}$

Hence the result.

■

এর উল্টো দিকটাও কিন্তু সত্যি, মানে যদি দ্যাখো যে, $\text{corr}(X,Y) = 1$, তবে নিশ্চিত জানবে যে এমন a, b, c আছে যাতে $aX + bY + c = 0$ হবে, যেখানে a, b-র চিহ্ন পরস্পরের বিপরীত। একইভাবে, $\text{corr}(X,Y) = -1$ হওয়া মানেও a, b, c থাকবে যাতে $aX + bY + c = 0$ হয়, কিন্তু এবার a, b-র একই চিহ্ন। আমরা এই উল্টোদিকটা এখানে প্রমাণ করব না। এবার দেখি কখন $\text{corr}(X,Y) = 0$ হয়। নীচের অংকটা সেই নিয়েই। যেহেতু $\text{corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$, তাই $\text{corr}(X,Y) = 0$ হওয়া মানে $\text{cov}(X,Y) = 0$. আর তার মানে হল $E(XY) = E(X)E(Y)$, কারণ $\text{cov}(X,Y) = E(XY) - E(X)E(Y)$.

Example 19: If X and Y are independent random variables, show that $E(XY) = E(X)E(Y)$ for both discrete and continuous cases. **(2008.3b)**

SOLUTION:

Discrete case: Let X, Y take values $\{x_1, x_2, ...\}$ and $\{y_1, y_2, ...\}$, respectively.
Since they are independent,
$\therefore \forall i, j \quad P(X = x_i,\ Y = y_j) = P(X = x_i)P(Y = y_j)$.
Now,

$$\begin{aligned} E(XY) &= \sum_i \sum_j x_i y_j P(X = x_i,\ Y = y_j) \\ &= \sum_i \sum_j x_i y_j P(X = x_i)P(Y = y_j) \quad [\because \text{independent}] \\ &= \left[\sum_i x_i P(X = x_i)\right] \times \left[\sum_j y_j P(Y = y_j)\right] \end{aligned}$$

যাবতীয় i-ওয়ালা জিনিসগুলোকে একজায়গায়, আর
যাবতীয় j-ওয়ালাদের এক জায়গায় করে এটা পেলাম।

$$= E(X)E(Y),$$

as required.

এবার continuous কেসটা করতে হবে। এখানে আমরা ধরে নেব যে PDF আছে।

Density case: Let X, Y have PDFs $f_X(x)$ and $f_Y(y)$, respectively.
Since they are independent, their joint PDF is

$$f(x,y) = f_X(x)f_Y(y).$$

Now,

$$\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dxdy \\
&= \left[\int_{-\infty}^{\infty} xf_X(x)dx\right] \times \left[\int_{-\infty}^{\infty} yf_Y(y)dy\right]
\end{aligned}$$

যাবতীয় x আর y-দের আলাদা করে লিখলেই এটা পাওয়া যায়।

$$= E(X)E(Y),$$

as required.

■

সুতরাং দেখা যাচ্ছে যে, X, Y যদি independent হয়, তবে ওদের covariance-টা শূন্য হবে, ফলে correlation-ও শূন্য হবে (যদি না $\text{var}(X)$ বা $\text{var}(Y)$-টা শূন্য হয়ে যায়, সেক্ষেত্রে correlation হবে undefined)।
এর উল্টো দিকটা কিন্তু সত্যি নয়। যদি দ্যাখো $\text{cov}(X,Y) = 0$ তা থেকেই সিদ্ধান্ত করা যায় না যে, X, Y পরস্পরের সঙ্গে independent হবেই। নীচের অংকটা করলে সে কথাটা ভালো করে মনে থাকবে।

Exercise 3: দুটো random variable দিচ্ছি X আর Y. তোমার কাজ হবে $\text{cov}(X,Y)$ বার করা। X খালি তিনটে value নিতে পারে, $-1, 0$ আর 1, প্রত্যেকটারই probability হল $\frac{1}{3}$ করে। আর $Y = X^2$. দ্যাখো তো $\text{cov}(X,Y)$ কত হয়। এখানে কি X আর Y-কে পরস্পরের সঙ্গে independent বলা যায়? ■

24.2.2 প্রয়োগ

এতক্ষণ যা যা ফর্মুলা শিখলাম, এবার তাদের কিছু প্রয়োগ দেখব। এই অংকগুলো নিছক খানিকটা কসরতমাত্র।

Example 20: Let $U = X + aY$, $V = X + \frac{\sigma_x}{\sigma_y}Y$, where a is a constant and σ_x, σ_y are the standard deviations of X, Y, where X, Y are positively correlated. If $\rho(U,V) = 0$, then show that $a = -\frac{\sigma_x}{\sigma_y}$.[5] **(2011.3d)**

SOLUTION: এখানে $\rho(U,V)$ মানে হল $\text{corr}(U,V)$.

Since $\rho(U,V)=0$, so $\operatorname{cov}(U,V)=0$.

$$\begin{aligned}\operatorname{cov}(U,V) &= \operatorname{cov}\left(X+aY, X+\frac{\sigma_x}{\sigma_y}Y\right)\\ &= \operatorname{var}(X)+a\frac{\sigma_x}{\sigma_y}\operatorname{var}(Y)+\left(a+\frac{\sigma_x}{\sigma_y}\right)\operatorname{cov}(X,Y)\end{aligned}$$

সেই যে বলেছিলাম covariance অনেকটা গুণ করার মত আচরণ করে, এখানে সেটা কাজে লাগল।

$$\begin{aligned} &= \sigma_x^2+a\frac{\sigma_x}{\sigma_y}\sigma_y^2+\left(a+\frac{\sigma_x}{\sigma_y}\right)\operatorname{cov}(X,Y)\\ &= \sigma_x^2+a\sigma_x\sigma_y+\left(a+\frac{\sigma_x}{\sigma_y}\right)\operatorname{cov}(X,Y)=0.\end{aligned}$$

এবার আমরা যাবতীয় a-ওয়ালা জিনিসকে বাঁদিকে রাখব--

So

$$a(\sigma_x\sigma_y+\operatorname{cov}(X,Y)) = -\sigma_x^2-\frac{\sigma_x}{\sigma_y}\operatorname{cov}(X,Y) = -\frac{\sigma_x}{\sigma_y}(\sigma_x\sigma_y+\operatorname{cov}(X,Y)).$$

Since $\operatorname{corr}(X,Y)>0$, so $\operatorname{cov}(X,Y)>0$.
So $\sigma_x\sigma_y+\operatorname{cov}(X,Y)\neq 0$, since $\sigma_x,\sigma_y>0$.
Hence, dividing by $\sigma_x\sigma_y+\operatorname{cov}(X,Y)$, we get $a=-\frac{\sigma_x}{\sigma_y}$, as required.

■

এবার একই রকম আরেকটা অংক।

Example 21: If X and Y are two correlated random variables with same standard deviation, show that the correlation coefficient between X and $X+Y$ is $\sqrt{(1+\rho)/2}$, where ρ is the correlation coefficient between X and Y.[5] **(2010.3b)**

SOLUTION: আমাদের $\operatorname{corr}(X,X+Y)$ বার করে দেখাতে হবে। তার জন্য দরকার $\operatorname{cov}(X,X+Y)$ আর $\operatorname{var}(X+Y)$. সেগুলো আগে বার করে নিই।

Let $\operatorname{var}(X)=\operatorname{var}(Y)=\sigma^2$.
Then $\operatorname{cov}(X,Y)=\rho\sigma^2$.
So

$$\begin{aligned}\operatorname{cov}(X,X+Y) &= \operatorname{var}(X)+\operatorname{cov}(X,Y)\\ &= \sigma^2+\rho\sigma^2\\ &= \sigma^2(1+\rho).\end{aligned}$$

Also

$$\begin{aligned}\mathrm{var}(X+Y) &= \mathrm{var}(X)+\mathrm{var}(Y)+2\mathrm{cov}(X,Y)\\ &= \sigma^2+\sigma^2+2\rho\sigma^2\\ &= 2\sigma^2(1+\rho).\end{aligned}$$

Hence

$$\mathrm{corr}(X,X+Y)=\frac{\mathrm{cov}(X,X+Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(X+Y)}}=\frac{\sigma^2(1+\rho)}{\sqrt{\sigma^2\times 2\sigma^2(1+\rho)}}=\sqrt{\frac{1+\rho}{2}},$$

as required.

■

Example 22: If the random variables $X_1,...,X_{2n}$ all have the same variance s^2 and the correlation coefficient between each pair (X_i,X_j) $i\neq j$, of them is ρ, show that the correlation coefficient between $\sum_{i=1}^n X_i$ and $\sum_{i=n+1}^{2n} X_i$ is

$$\frac{n\rho}{1+(n-1)\rho}.$$

[8] **(2004.5a)**
SOLUTION: একটা correlation বার করতে হবে। সুতরাং একটা covariance আর দুটো variance বার করে নিতে হবে আগে।

For $i\neq j$, we have $\mathrm{cov}(X_i,X_j)=\mathrm{corr}(X_i,X_j)\times\sqrt{\mathrm{var}(X_i)\mathrm{var}(X_j)}=\rho s^2$.

$$\begin{aligned}\mathrm{cov}\left(\sum_{i=1}^n X_i, \sum_{i=n+1}^{2n} X_i\right) &= \sum_{i=1}^n\sum_{j=n+1}^{2n}\mathrm{cov}(X_i,X_j)\\ &= \sum_{i=1}^n\sum_{j=n+1}^{2n}\rho s^2\\ &= n^2\rho s^2.\end{aligned}$$

এখানে কেন n^2 হল, বুঝলে তো? কারণ, i নিতে পারে n-খানা value, তার প্রতিটার জন্য j-ও নিতে পারে ঠিক n-খানা value. তাই $n\times n$ হয়ে n^2 হয়েছে।

Also

$$\begin{aligned}\operatorname{var}\left(\sum_{i=1}^n X_i\right) &= \sum_{i=1}^n \operatorname{var}(X_i) + \sum_{i\neq j}^n \operatorname{cov}(X_i, X_j)\\ &= \sum_{i=1}^n s^2 + \sum_{i\neq j}^n \rho s^2\\ &= ns^2 + n(n-1)\rho s^2 = ns^2(1+(n-1)\rho).\end{aligned}$$

এখানে $n(n-1)$ হয়েছে কারণ আগে যেখানে n^2 এসেছিল, এখানেও সেরকমই, খালি $i \neq j$ হওয়ায় n-খানা কেস বাদ গেছে। তাই $n^2 - n = n(n-1)$ হয়েছে।

Similarly,

$$\operatorname{var}\left(\sum_{i=n+1}^{2n} X_i\right) = ns^2(1+(n-1)\rho).$$

এইবার তবে correlation-এ হাত দেওয়া যাক। এর জন্য ধরে নিতে হবে যে variance-গুলো 0 নয়। সেটা অংকে বলে দেওয়া নেই। তাই আমরাই আলাদা করে উল্লেখ করে নেব।

We shall assume $1+(n-1)\rho \neq 0$, otherwise the required correlation will not be defined.

So

$$\begin{aligned}\operatorname{corr}\left(\sum_{i=1}^n X_i, \sum_{i=n+1}^{2n} X_i\right) &= \frac{\operatorname{cov}\left(\sum_{i=1}^n X_i, \sum_{i=n+1}^{2n} X_i\right)}{\sqrt{\operatorname{var}\left(\sum_{i=1}^n X_i\right)\operatorname{var}\left(\sum_{i=n+1}^{2n} X_i\right)}}\\ &= \frac{n^2\rho s^2}{ns^2 + n(n-1)\rho s^2}\\ &= \frac{n\rho}{1+(n-1)\rho},\end{aligned}$$

as required.

■

DAY 25
Bivariate distributions (part 4)

25.1 Regression

কয়েকটা উদাহরণ দিয়ে প্রথমে ধারণাটা বুঝে নিই।

Example 23: ধরো একটা unbiased কয়েন নিয়ে 10 বার টস করব। তোমাকে আগেভাগে আন্দাজ করতে হবে ঠিক

কটা head পড়বে। এরকম ক্ষেত্রে একেবারে নিশ্চিত হয়ে তো আর বলা যায় না, কিন্তু একটা ভালো উত্তর হল $10 \times \frac{1}{2} = 5$ বলা। কারণ যদি X হয় head-এর সংখ্যা, তবে $E(X) = 5$. সেই অর্থে $E(X)$ যেন X-এর সম্ভাব্য যাবতীয় value-র ভালো একটা প্রতিনিধি। ■

এবার একটু কঠিন উদাহরণ। এই উদাহরণের একটা সংস্করণ আমরা আগেই দেখেছি conditional distribution শেখার সময়ে।

Example 24: প্রথমে একটা ছক্কা চাললাম। যত সংখ্যা আসবে, তত গুণ দশবার একটা কয়েন টস করা হবে। যেমন ছক্কায় যদি 2 পড়ে, তবে $2 \times 10 = 20$ বার টস করব। ছক্কা আর কয়েন, দুটোই unbiased. এবার ধরো তোমাকে বলে দিলাম ছক্কায় 3 পড়েছে। এটুকু মাত্র শুনে আন্দাজ করতে হবে কতগুলো head পড়বে। যেহেতু ছক্কায় 3 পড়েছে, সুতরাং টস করা হবে $3 \times 10 = 30$ বার। সুতরাং একটা ভালো আন্দাজ হল $30 \times \frac{1}{2} = 15$. ■

এখানে কাজটা কী করলাম দেখি। ধরো ছক্কার outcome-টাকে বললাম X আর মোট head-এর সংখ্যাকে বললাম Y. তাহলে $X = 3$ বলে দিয়ে আন্দাজ করতে বলা হয়েছিল Y কত হতে যাচ্ছে। তার জন্য প্রথমে আমরা $X = 3$ হলে Y-এর conditional distribution-টা দেখলাম (যেটা হল Binomial $\left(30, \frac{1}{2}\right)$), এবং তার expectation-টা বার করলাম, মানে $E(Y|X = 3)$. একইভাবে যদি X-এর অন্য কোনো value (ধরো x) দেওয়া থাকত তবে উত্তরটা হত $E(Y|X = x)$. আমাদের উদাহরণে $X = x$ হলে কয়েনটা $10x$ বার টস করা হবে, সুতরাং $E(Y|X = x) = 10x \times \frac{1}{2} = 5x$ হবে। লক্ষ কর যে, এটা x-এর একটা function. একে বলে **regression function of Y on X**. একইভাবে, যদি $Y = y$ বলে দিয়ে X-কে আন্দাজ করতে বলা হত তবে ব্যবহার করতাম $E(X|Y = y)$, যেটা হত y-এর একটা function. সেই function-টাকে বলে **regression function of X on Y**. অনেক সময়ে এদেরকে regression function না বলে **regression curve**-ও বলা হয়।
একটা ছোট্টো অংক করে ব্যাপারটা সড়গড় হয়ে নিই।

Example 25: The joint probability density function of two random variables, X and Y, is given by

$$f(x, y) = \frac{1}{\pi\sqrt{3}} e^{-2\left(x^2 - xy + \frac{y^2}{3}\right)}.$$

Find the regression curve of Y on X.[5] **(2012.3d)**

SOLUTION:

> The regression curve of Y on X is given by
>
> $$E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy,$$
>
> where $f_{Y|X}(y|x)$ is the conditional density of Y given $X = x$.

Now

$$f_{Y|X}(y|x) = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dy}$$

দেখে মনে হচ্ছে যেন এবার নীচের integral-টা বার করতে হবে।
কিন্তু আসলে তার দরকার নেই। আমরা দেখব কোথায় কোথায় y আছে।
লক্ষ কর যে, নীচের তলায় কিন্তু কোনো y নেই,
যেহেতু y-এর সাপেক্ষে definite integration করা হয়েছে।
তাই আমরা y-ছাড়া সব জিনিসকে পোঁটলা করে এভাবে লিখতে পারি--

$$= (\text{terms free of } y) \times e^{-2\left(x^2-xy+\frac{y^2}{3}\right)}$$

এখানে $e^{-2\left(x^2-xy+\frac{y^2}{3}\right)} = e^{-2x^2} \times e^{-2\left(-xy+\frac{y^2}{3}\right)}$.
যেহেতু e^{-2x^2}-এর মধ্যে কোনো y নেই, তাই ওটাকেও
পোঁটলার ভিতর ঢুকিয়ে নেব--

$$= (\text{terms free of } y) \times e^{-2\left(-xy+\frac{y^2}{3}\right)}$$

$$= (\text{terms free of } y) \times e^{-\frac{2}{3}(y^2-3xy)}$$

এবার একটা square complete করব, $y^2 - 3xy = \left(y - \frac{3}{2}x\right)^2 - \frac{9}{4}x^2$.
ওই যে, $-\frac{9}{4}x^2$-টা আমদানি হল, ওটাও y-ছাড়া,
তাই ওকেও পোঁটলার মধ্যে চালান করে দেওয়া যাবে।

$$= (\text{terms free of } y) \times e^{-\frac{2}{3}\left(y-\frac{3}{2}x\right)^2}$$

এইবার লক্ষ কর যে, পুরো জিনিসটা একটা normal distribution-এর PDF-এর
মত চেহারা নিতে যাচ্ছে। সেই সাদৃশ্যটাকে আরও স্পষ্ট করে তুলতে আমরা
এইভাবে সাজিয়ে লিখব--

$$= (\text{terms free of } y) \times e^{-\frac{1}{2\times\frac{3}{4}}\left(y-\frac{3x}{2}\right)^2}.$$

এইবার লক্ষ কর যে, PDF-টা দেখতে হয়েছে এইরকম--

$$\text{constant} \times e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

এই চেহারাটা আমাদের পরিচিত, সুতরাং ওই constant-টা কী সেটা না বার করেই বলে দিতে পারি যে, এটা $N(\mu,\sigma)$-র PDF.

Comparing with $N(\mu,\sigma)$ density we see that the conditional density of Y given $X=x$ is normal with mean $\frac{3x}{2}$ and variance $\frac{3}{4}$.

Thus $E(Y|X=x) = \frac{3x}{2}$.

So required regression curve of Y on X is $f(x) = \frac{3x}{2}$.

■

25.1.1 Least squares lines

আমরা regression curve বার করা শিখলাম। যদি X, Y দুটো random variable হয়, যারা jointly distributed, তাহলে "regression curve of Y on X"-কে মনে করতে পারো যেন X-এর কোনো value বলে দিলে তার ভিত্তিতে Y-এর value আন্দাজ করার একটা ভালো কায়দা[2]। এই regression curve-টা হবে x-এর একটা function, ধরো $h(x)$. যদি বলে দিই যে $X = 5$ হয়েছে, তাহলে Y-এর value-কে তুমি আন্দাজ করবে $h(5)$, এইরকম। বুঝতেই পারছ যে, যদি $h(x)$-এর ফর্মুলাটা খুব জটিল কিছু হয়, তবে কাজটা বেশ কঠিন হবে। সেই কারণে লোকে এর একটা সহজ সংস্করণ ভেবে রেখেছে, যেখানে $h(x)$ একটা সরলরেখা হবে, মানে $h(x) = \alpha + \beta x$ জাতীয় হবে। ব্যাপারটা এইরকম--

> তোমাকে $X = x$ বলে দেওয়া হবে, তার ভিত্তিতে তোমাকে $\alpha + \beta x$ জাতীয় ফর্মুলাই ব্যবহার করে আন্দাজ করতে হবে Y-এর value. কীভাবে তুমি α, β নির্বাচন করবে যাতে আন্দাজটায় ভুলের পরিমাণ যথাসম্ভব কম থাকে?

এর একটা কায়দা হল এমনভাবে α, β নেওয়া যাতে $E\big((Y - \alpha - \beta X)^2\big)$ যথাসম্ভব কম হয়। এইটা কোথা থেকে পেলাম বুঝতে পারছ নিশ্চয়ই? Y-কে আন্দাজ করছি $\alpha + \beta X$ দিয়ে, তাই ভুলের পরিমাণ হল $Y - \alpha - \beta X$. সেটা positive নাকি negative, তা নিয়ে আমাদের চিন্তা নেই, চিন্তা খালি ভুলের পরিমাণটা নিয়ে, তাই square করেছি, $(Y - \alpha - \beta X)^2$. এটা নিজেই একটা random variable, তাই কখনও বড় হবে কখনো ছোটো, তার উপরে আমাদের পূর্ণ নিয়ন্ত্রণ নেই, সেই কারণে expectation নিয়েছি, মানে মোটামুটিভাবে কত value হয়, সেটাকে minimise করব। অর্থাৎ $E\big((Y - \alpha - \beta X)^2\big)$-কে minimise করব। লক্ষ কর যে, square-টাকে বড় করে লিখে expectation-এর linearity ব্যবহার করলে হয়--

$$E\big((Y - \alpha - \beta X)^2\big) = E(Y^2) + \alpha^2 + \beta^2 E(X^2) - 2aE(Y) - 2bE(XY) + 2abE(X).$$

এবার যদি α আর β-এর সাপেক্ষে এটাকে differentiate করে, শূন্যের সমান করে দাও, তবে খানিকটা অংক কষার পর উত্তর পাবে এইটা--

$$\alpha = E(Y) - \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)} E(X) \text{ আর } \beta = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)}.$$

এই দুটো ফর্মুলার চেহারা আমাদের পূর্বপরিচিত। Least square দিয়ে সরলরেখা ফিট করার সময়ে আমরা ঠিক এইরকম জিনিসই পেয়েছিলাম, খালি সেখানে covariance আর variance ছিল sample-এর ভিত্তিতে বার করা, এবং $E(X)$ আর $E(Y)$-এর জায়গায় ছিল, যথাক্রমে, $\bar{x}$ এবং $\bar{y}$. এই α, β বসানোর পর $Y = \alpha + \beta X$-এর চেহারা হবে--

$$Y - E(Y) = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)} (X - E(X)).$$

একে বলে **least square regression line of Y on X**. এখানে slope-টাকে বলে **regression coefficient of Y on X**, আর লেখে $\beta_{Y|X}$, মানে

$$\beta_{Y|X} = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)}.$$

যদি X আর Y-এর স্থানবিনিময় করো (মানে এবার Y দেওয়া আছে, আর তার ভিত্তিতে X আন্দাজ করছি), তবে ফর্মুলাটা হবে

$$X - E(X) = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(Y)} (Y - E(Y)).$$

এখান থেকে পাব **regression coefficient of X on Y**,

$$\beta_{X|Y} = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(Y)}.$$

Example 26: Write down the least square regression lines of Y on X and of X on Y. **(2008.3c)**

SOLUTION:

[2] একইভাবে X, Y-এর স্থান বিনিময় করলে regression curve of X on Y-ও পাওয়া যায়।

> The least square regression line of Y on X is
> $$Y = E(Y) + \beta_{Y|X}(X - E(X)),$$
> where $\beta_{Y|X} = \frac{\text{cov}(X,Y)}{\text{var}(X)}$.
> Similarly, the least square regression line of X on Y is
> $$x = E(X) + \beta_{X|Y}(y - E(Y)),$$
> where $\beta_{X|Y} = \frac{\text{cov}(X,Y)}{\text{var}(Y)}$.

■

Example 27: The bivariate probability density function of two random variables X and Y is given by
$$f_{x,y}(x,y) = 3x^2 - 8xy + 6y^2, \quad 0 < x < 1, \quad 0 < y < 1.$$
Find the regression curves for the means and also the least square regression lines.[5] **(2014.4b)**
SOLUTION: আমরা যাকে regression curve বলেছি, সেটাকেই বড় করে লিখেছে regression curve for the means.
আমরা জানি যে, regression curve বার করা মানে conditional expectation বার করা। তার জন্য চাই conditional PDF, এবং তার জন্য আবার চাই marginal PDF দুটো। প্রথমে marginal PDF দুটো বার করে নিই।

> The marginal PDF of X is
> $$f_X(x) = \int_{-\infty}^{\infty} f_{x,y}(x,y)dy = \cdots = \begin{cases} 3x^2 - 4x + 2 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

মাঝের মামূলী কয়েকটা ধাপ এখানে আর দেখাইনি।

> The marginal PDF of Y is
> $$f_Y(y) = \int_{-\infty}^{\infty} f_{x,y}(x,y)dx = \cdots = \begin{cases} 6y^2 - 4y + 1 & \text{if } 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

এবার conditional PDF দুটোর পালা--

> Conditional density of Y given $X = x \in (0,1)$ is
> $$f_{Y|X}(y|x) = \frac{f_{x,y}(x,y)}{f_X(x)} = \begin{cases} \frac{3x^2-8xy+6y^2}{3x^2-4x+2} & \text{if } 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}.$$

> Conditional density of X given $Y = y \in (0,1)$ is
>
> $$f_{X|Y}(x|y) = \frac{f_{x,y}(x,y)}{f_y(y)} = \begin{cases} \frac{3x^2-8xy+6y^2}{6y^2-4y+1} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

এবার অবশেষে conditional expectation দুটো বার করতে পারব--

> Regression curve of Y on X is
>
> $$E(Y|X=x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy = \frac{\int_0^1 y(3x^2-8xy+6y^2)dy}{3x^2-4x+2} = \cdots = \frac{9x^2-16x+9}{6(3x^2-4x+2)}.$$

এখানে কিন্তু integration-টা আসলে খুবই সহজ। যেহেতু y-এর সাপেক্ষে integration হচ্ছে, তাই x-কে constant বলে ভাবতে পারো, সুতরাং denominator-টা পুরোই integration-এর বাইরে বেরিয়ে আসবে।

> Similarly, the regression curve of X on Y is
>
> $$E(X|Y=y) = \int_{-\infty}^{\infty} x f_{x|y}(x|y)dx = \frac{\int_0^1 x(3x^2-8xy+6y^2)dx}{6y^2-4y+1} = \frac{36y^2-32y+9}{12(6y^2-4y+1)}.$$

অনেকগুলো integration টপাটপ কষে ফেলা গেল। কিন্তু ব্যাপারটা ঠিক কী হচ্ছে সেটা হয়তো বোঝা যাচ্ছে না। একটু ছবি এঁকে সেটা বুঝে নেওয়া যাক। আমরা যে joint PDF-টা নিয়ে শুরু করেছিলাম, সেই $f_{x,y}(x,y)$-এর গ্রাফ রয়েছে Fig 33-এ। কতকটা রুমালের মত দেখতে, একটা প্রান্ত বেশ অনেকটা উঠে আছে, তার বিপরীত প্রান্তটাও খানিকটা উঠে আছে, মাঝখানটা মাটিতে প্রায় লাগা। মনে রেখো যে bivariate PDF মানেই এর পিছনে এক জোড়া random variable আছে, X, Y, এবং তাদের পিছনে আছে একটা random experiment. এই random experiment-টা সম্বন্ধে একটা ধারণা থাকলে ব্যাপারটা বুঝতে সুবিধা হবে। সেই জন্য experiment-টাকে আমরা বারবার করে করেছি কম্পিউটারে, ফলে অনেকগুলো $(x_1, y_1), ..., (x_n, y_n)$ পেয়েছি। তাদের scatterplot এঁকেছি Fig 34-এ। রুমালের যে প্রান্তটা বেশী উঠে ছিল, সেখানেই point-গুলোর বেশী ভীড় করে আছে। সেটাই স্বাভাবিক, কারণ density মানে তো ঘনত্ব, যেখানে density বেশী ভীড়ের ঘনত্বও সেখানেই বেশী। রুমালের অন্যপ্রান্তটাও একটু উঠেছিল, বলে ওখানেও কিছু point-এর সামান্য জটলা হয়েছে, তবে তারা আছে ছড়িয়ে ছিটিয়ে। এবার regression curve of Y on X-এর কাজ হল এমন একটা সম্পর্ক বার করা, যা দিয়ে X জানা থাকলে Y-এর মোটামুটি একটা value আন্দাজ করা যাবে। আমরা অংক কষে যে $E(Y|X=x)$-এর ফর্মুলা বার করেছিলাম সেটার গ্রাফ দেখিয়েছি Fig 35-এ। লক্ষ কর গ্রাফটা কীরকম বাঁদিকের ভীড় থেকে বেরোনোর পর বেঁকে গিয়ে ছড়ানো ছেটানো point-গুলোর মধ্য দিয়ে গেছে।
এবার আমরা least square লাইনদুটো বার করব। এখানেও উদ্দেশ্য একই--X, Y-এর মধ্যে একটা দেওয়া থাকলে অন্যটার একটা মোটামুটি value আন্দাজ করতে পারা, খালি এবার কাজটা করতে হবে সরলরেখা দিয়ে। তাই একইসঙ্গে ভীড়টাকে এবং ছড়ানো ছেটানো point-গুলোকে সামলানো শক্ত।

Fig 33

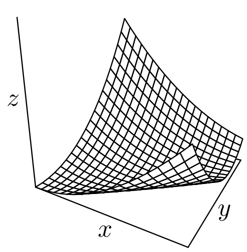


Fig 34

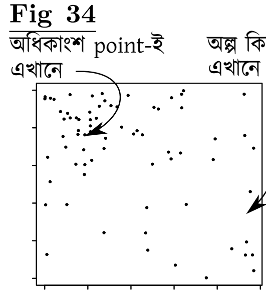
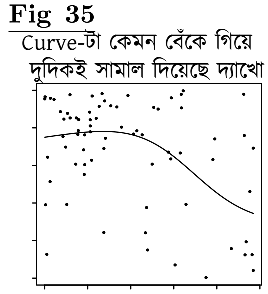


Fig 35



আমরা least square line-এর চেহারা তো জানিই। ওদের জন্য যা যা উপকরণ দরকার সেগুলোকে এক এক করে বার করে ফেলি--

Least square lines:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_x(x)dx = \tfrac{5}{12} \\ E(Y) &= \int_{-\infty}^{\infty} y f_y(y)dy = \tfrac{2}{3} \\ E(X^2) &= \int_{-\infty}^{\infty} x^2 f_x(x)dx = \tfrac{4}{15} \\ E(Y^2) &= \int_{-\infty}^{\infty} y^2 f_y(y)dy = \tfrac{8}{15} \\ E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{x,y}(x,y)dxdy = \tfrac{1}{3}. \end{aligned}$$

Thus

$$\begin{aligned} \mathrm{var}(X) &= E(X^2) - (E(X))^2 = \tfrac{4}{15} - \left(\tfrac{5}{12}\right)^2 = \tfrac{67}{720} \\ \mathrm{var}(Y) &= E(Y^2) - (E(Y))^2 = \tfrac{8}{15} - \left(\tfrac{2}{3}\right)^2 = \tfrac{4}{45} \\ \mathrm{cov}(X,Y) &= E(XY) - E(X)E(Y) = \tfrac{1}{3} - \tfrac{5}{12} \times \tfrac{2}{3} = \tfrac{1}{18} \end{aligned}$$

So the least square regression line of Y on X is

$$(y - E(Y)) = \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)}(x - E(X)),$$

or

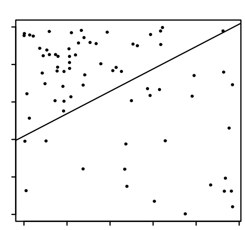
$$(y - \tfrac{2}{3}) = \frac{\frac{1}{18}}{\frac{67}{720}}(x - \tfrac{5}{12}),$$

or

$$y = \tfrac{28}{67} + \tfrac{40}{67}x.$$

ছবিটা রয়েছে Fig 36-এ। এখানে ছড়ানো ছেটানো point-গুলোর দিকে মন দিতে গিয়ে লাইনটা ভীড়ের থেকে একটু বেরিয়ে পড়েছে।

Fig 36



> Similarly the least square regression line of X on Y is $(x - E(X)) = \frac{\text{cov}(X,Y)}{\text{var}(Y)}(y - E(Y))$, ie,
> $$x = \frac{5}{8}y.$$

■

Exercise 4: The bivariate pdf of two random variables X and Y is given by

$$f(x, y) = \begin{cases} x + y & \text{if } 0 < x < 1, \quad 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Calculate the means, standard deviations of X and Y, and also the correlation coefficient between X and Y. Find the equations of the two regression lines.[10] **(2005.5a)**

HINT:
আগের অংকের অনুকরণে কর। তবে এখানে কাজ অনেকটাই সোজা, কারণ PDF-টা x, y-এর সাপেক্ষে symmetric, অর্থাৎ x, y-এর স্থান বিনিময় করলেও PDF-টা বদলাবে না। তাই $E(X) = E(Y), \text{var}(X) = \text{var}(Y)$ হবে। একই কারণে খালি regression line of Y on X বার করলেই সেখান থেকে x, y-এর স্থান বিনিময় করে regression line of X on Y-ও পাওয়া যাবে। ■

এবার least square regression-এর কিছু ধর্ম শেখা যাক।
একটু চিন্তা করলেই নীচের theorem-টা পেয়ে যাবে--

> **THEOREM**
>
> If X, Y are two jointly distributed random variables, then $\beta_{X|Y}, \beta_{Y|X}, \text{cov}(X, Y), \text{corr}(X, Y)$ all have the same sign, and
> $$\text{corr}(X, Y)^2 = \beta_{X|Y}\beta_{Y|X}.$$

এর পরের theorem-টাও প্রমাণ করা খুবই সহজ--

> **THEOREM**
>
> The two regression lines for two jointly distributed random variables X, Y are the same if and only if either $\text{corr}(X, Y) = 1$ or $\text{corr}(X, Y) = -1$. Otherwise, they intersect at the point $\big(E(X), E(Y)\big)$.

নীচের অংকে আরেকটা ধর্ম আলোচনা করা হয়েছে। কিন্তু সেই ধর্মটা বাস্তবে বড় একটা কাজে আসে না।

Example 28: Write down the least square regression lines of Y

on X and of X on Y. Show that the acute angle between these lines is

$$\tan^{-1}\left(\frac{1-\rho^2}{\rho}\times\frac{\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\right)$$

(the symbols have their usual meanings). What happens when $\rho = \pm 1$? **(2008.3c)**
SOLUTION: প্রথম অংশ তো আগেই করেছি। তাও একবার লিখে নিই, যাতে দ্বিতীয় অংশটা বুঝতে সুবিধা হয়।

> The least square regression line of Y on X is
>
> $$Y = \mu_y + \beta_{Y|X}(X-\mu_x),$$
>
> where $\beta_{y|x} = \frac{\rho\sigma_y}{\sigma_x}$.
> Similarly, the least square regression line of X on Y is
>
> $$X = \mu_x + \beta_{X|Y}(Y-\mu_y),$$
>
> where $\beta_{X|Y} = \frac{\rho\sigma_x}{\sigma_y}$.

এখানে একটু coordinate geometry মনে করে নিতে হবে।

> We know that the acute angle between two intersecting straight lines with slopes m_1, m_2 is
>
> $$\tan^{-1}\left(\frac{|m_1-m_2|}{1+m_1m_2}\right).$$

এখানে absolute value-টা লক্ষ কর। ওটা আছে বলেই আমরা acute angle পাচ্ছি।

> Here
>
> $$m_1 = \beta_{Y|X} = \frac{\rho\sigma_y}{\sigma_x} \text{ and } m_2 = 1/\beta_{X|Y} = \frac{\sigma_y}{\rho\sigma_x}.$$

এখানে $m_2 = 1/\beta_{X|Y}$ কেন হল বুঝলে তো? যখন X on Y লাইনটা বার করছি তখন আমরা X-কে রাখছি vertical দিকে আর Y-কে horizontal দিকে। তাতে slope-টা ছিল $\beta_{X|Y}$. কিন্তু acute angle-এর ফর্মুলাতে গ্রাফ আঁকার সাধারণ নিয়ম ব্যবহার করা হয়, মানে Y থাকে vertical দিকে, X থাকে horizontal দিকে। সেই কারণে X, Y-এর স্থানবিনিময়ের দরকার পড়ে, ফলে $\beta_{X|Y}$-টা উল্টে $1/\beta_{X|Y}$ হয়ে গেছে।
লক্ষ কর যে, ρ, σ_x, σ_y প্রত্যেকেই অনেক সময়ে denominator-এ থাকছে। তাই ওরা যে nonzero সেটা ধরে নিতে হবে।

> Here we assume $\rho \neq 0$ and $\sigma_x, \sigma_y > 0$.

যদি $\rho = 0$ হত, তবে দুটো regression লাইন পরস্পরের সঙ্গে $90°$ কোণে থাকত, তাই ওদের মধ্যে কোনো acute angle থাকতই না। আমরা সেই কেসটা নিয়ে এখানে মাথা ঘামাব না।

So

$$
\begin{aligned}
\frac{|m_1 - m_2|}{1 + m_1 m_2} &= \frac{|[|\frac{\rho\sigma_y}{\sigma_x} - \frac{\sigma_y}{\rho\sigma_x}|]|}{1 + \frac{\sigma_y^2}{\sigma_x^2}} \\
&= \frac{|[|\rho\sigma_x\sigma_y - \frac{\sigma_x\sigma_y}{\rho}|]|}{\sigma_x^2 + \sigma_y^2}
\end{aligned}
$$

এটা পেলাম উপর নীচ দুদিককেই σ_x^2 দিয়ে গুণ করে।

$$
\begin{aligned}
&= \sigma_x\sigma_y \frac{|[|\rho - \frac{1}{\rho}|]|}{\sigma_x^2 + \sigma_y^2} \\
&= \frac{|\rho^2 - 1|}{\rho} \times \frac{\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \\
&= \frac{1 - \rho^2}{\rho} \times \frac{\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \quad [\because \rho^2 \leq 1],
\end{aligned}
$$

as required.

এবার দ্বিতীয় অংশ, যেখানে দেখতে হবে $\rho = -1$ বা $\rho = 1$ হলে কী হয়।

When $\rho = \pm 1$, then $1 - \rho^2 = 0$, so the acute angle becomes $\tan^{-1} 0 = 0$.
Thus, in this case, the two lines are the same.

■

Example 29: For two random variables X and Y with the same mean, the two regression lines are $y = ax + b$ and $x = \alpha y + \beta$. Show that

$$\frac{b}{\beta} = \frac{1 - a}{1 - \alpha}.$$

Find also the common mean.[5] **(2005.5b)**

SOLUTION:

Let the common mean be m, ie, $E(X) = E(Y) = m$.
We know that both the regression lines pass through $\big(E(X), E(Y)\big) = (m, m)$.
So

$$
\begin{aligned}
m &= am + b, \\
m &= \alpha m + \beta.
\end{aligned}
$$

Thus

$$m = \frac{b}{1 - a} = \frac{\beta}{1 - \alpha}. \qquad (*)$$

> Hence.
> $$\frac{b}{\beta} = \frac{1-a}{1-\alpha},$$
> as required.

এই যে আমরা ফস্ করে $\beta, 1-a$ আর $1-\alpha$-কে denominator-এ পাঠিয়ে দিলাম, এতে মনে একটু অস্বস্তি হওয়া স্বাভাবিক, কারণ ওরা যে nonzero এটা কী করে জানলাম? সেটা আমরা সত্যিই জানিনা। আসলে অংকটা পুরো ঠিক নয়, এখানে বলে দেওয়া উচিত ছিল যে $\beta \neq 0$ এবং $a, \alpha \neq 1$. তা, বলে যখন দেয়নি, আমরাই সেটা ছোট্টো করে বলে নিই--

> Here we assume $\beta \neq 0$ and $a, \alpha \neq 1$.

এবার common mean-টা, মানে m বার করতে বলেছে। সেটা তো (*)-এই বেরিয়ে গেছিল।

> Also the common mean is given by (*).

■

Answers

1. লক্ষ কর যে $r = \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum(X_i-\bar{X})^2\sum(Y_i-\bar{Y})}}$. যে কেসটা বাদ দিচ্ছি সেটাতে $(X=7, Y=9)$. এদিকে $\bar{X}=7$ এবং $\bar{Y}=9$. তাই এই কেসটার জন্য $(X_i-\bar{X}) = (Y_i-\bar{Y}) = 0$. তাই কিছুই বদলাবে না। **3.** $\text{cov}(X,Y)=0$. না, X, Y মোটেই পরস্পরের সঙ্গে independent নয়, কারণ $P(Y=0, X=1) = 0 \neq P(Y=0)P(X=1)$. **4.** Mean দুটোই $\frac{7}{12}$. দুটো standard deviation-ই হবে $\frac{\sqrt{11}}{12}$. আর correlation coefficient আসবে $-\frac{1}{11}$. Regression line of Y on X হবে $\left(y-\frac{7}{12}\right) = -\frac{1}{11}\left(x-\frac{7}{12}\right)$. অন্য regression line-টা হবে $\left(x-\frac{7}{12}\right) = -\frac{1}{11}\left(y-\frac{7}{12}\right)$.

Chapter VII
Limits

DAY 26
Chebyshev inequality

আমরা দেখেছি যে, X একটা random variable হলে সেটা বিভিন্ন value নিতে পারে, এবং $E(X)$ মোটামুটিভাবে এই value-গুলোর মাঝামাঝি জায়গায় থাকে (Fig 1)। ঠিক যেন $E(X)$ হল মা, আর X তার চঞ্চল বাচ্চা। মা-কে ঘিরে বাচ্চার লাফালাফি লেগেই থাকে। যদি বাচ্চাটা বেশী দুরন্ত হয়, তবে সে মাঝে মাঝেই মায়ের থেকে বেশী দূরে চলে যাবে, আবার শান্তশিষ্ট নির্জীবগোছের বাচ্চা হলে মায়ের আঁচল ধরেই একটু এদিক ওদিক করবে। কতটা চঞ্চল হলে কতদূর অব্ধি যাবার সম্ভাবনা, এইটা বেশ সুন্দর করে গুছিয়ে লেখার একটা কায়দা আছে, তাকে বলে Chebyshev inequality, যেটা এবার আমরা শিখব।

Fig 2-এ আমরা $E(X)$-এর দুপাশে সমান দূরত্বে দুটো বেড়া বসিয়েছি, দূরত্বটার নাম দিয়েছি ϵ. আমরা জানতে চাই বাচ্চাটার পক্ষে এই বেড়ার বাইরে বেরিয়ে পড়ার সম্ভাবনা কত, মানে $P(|X-E(X)|\geq\epsilon)$ কত? এখানে ϵ যেন মায়ের শাসনের পরিমাণ, কড়া মায়েরা বাচ্চাকে একটুও স্বাধীনতা দিতে চান না, তাঁদের বেলায় ϵ ছোটো। আবার কিছু মা আছেন, যাঁরা বাচ্চা অনেকটা দূরে গিয়ে খেললেও উতলা হন না, তাঁদের বেলায় ϵ বড়। এই বাচ্চার উপমার কথা মাথায় রেখে চট্ করে ভেবে নেওয়া যাক $P(|X-E(X)|\geq\epsilon)$ কখন বেশী বড় হবে, ϵ ছোটো হলে নাকি বড় হলে? যে সব মা পদে পদে বাধানিষেধ চাপান, বাধ্য হয়েই বাচ্চাকে সেই বাধানিষেধ ঘন ঘন অগ্রাহ্য করতে হয়। সুতরাং $\epsilon>0$ কমলে probability-টা বাড়বে। আবার একই ϵ-র জন্য চঞ্চল বাচ্চারা বেশী ঘনঘন বেড়া অতিক্রম করবে। একটা random variable-এর চাঞ্চল্য মাপার অন্যতম হাতিয়ার হল তার variance. সুতরাং অনুমান করতে পারছ যে, $\mathrm{var}(X)$ বাড়লে probability-টা বাড়ার কথা। এই দুটো জিনিসকে একসঙ্গে মেশালেই পাবে Chebyshev inequality–

Chebyshev inequality

Let X be a random variable with finite variance. Then

$$\forall\epsilon>0\quad P(|X-E(X)|\geq\epsilon)\leq\frac{\mathrm{var}(X)}{\epsilon^2}.$$

Fig 1

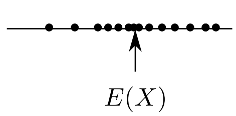
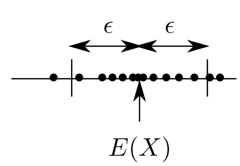


Fig 2



এরই প্রমাণ চেয়েছে নীচের অংকে।

Example 1: If X is any random variable having a finite variance, then prove that for any $\epsilon > 0$

$$P(|X - m| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2},$$

where m and σ, respectively, denote the mean and standard deviation of X.[5] **(2010.3c)**

SOLUTION:

Let

$$Y = \begin{cases} 1 & \text{if } |X - m| \geq \epsilon \\ 0 & \text{otherwise} \end{cases}$$

and $Z = 1 - Y$.

এই Y, Z-এর গল্পটা যদি বুঝতে অসুবিধা হয়, তবে একটা উদাহরণ দিয়ে ভাবো। মনে করো X হল একটা ছক্কা চালার outcome. ছক্কাটার সব দিক সমান ভারী, তাই $m = E(X) = (1 + \cdots + 6)/6 = 3.5$. ধরো $\epsilon = 2$ নিলাম। এবার তোমাকে দেখতে বললাম X কখন কখন $(m - \epsilon, m + \epsilon)$-এর বাইরে আছে। এর জন্য প্রথমে X-এর সব value এক লাইনে লিখে নেব, এবং যে যে value আমাদের শর্তটা পালন করবে (মানে $(3.5 - 2, 3.5 + 2) = (1.5, 5.5)$-এর বাইরে থাকবে) তাদের মাথায় একটা করে টিক চিহ্ন দেব, আর বাকিদের মাথায় কাটা চিহ্ন--

✓	×	×	×	×	✓
1	2	3	4	5	6

এবার যদি টিক চিহ্নটাকে 1 আর কাটা চিহ্নটাকে 0 ভাবো, তবেই Y পাবে। আর যদি কাটাকে 1 এবং টিককে 0 ধরো, তবে পাবে Z. লক্ষ কর যে, সর্বদাই $Y + Z = 1$ হচ্ছে। এইভাবে Y, Zদিয়ে লেখার সুবিধা হল এক ঢিলে discreteআর continuous দুটো ক্ষেত্রই হয়ে যায়।

এবার expectation-এর একটা ধর্মের কথা মনে করিয়ে দিই। যদি U, V দুটো random variable হয় যেখানে সব সময়েই $U \leq V$, তাহলে $E(U) \leq E(V)$ হবে। এখানে অবশ্যই ধরে নিচ্ছি যে, $E(U), E(V)$ দুজনেই exist করে। এই জিনিসটা এক্ষুণি কাজে লাগবে।

We know that

$$\begin{aligned} \sigma^2 &= E((X - m)^2) \\ &= E((X - m)^2(Y + Z)) \quad [\because Y + Z = 1] \end{aligned}$$

এই $Y + Z = 1$ দিয়ে গুণ করে লেখাটা একটা কৌশল।

$$\begin{aligned} &= E((X - m)^2 Y) + E((X - m)^2 Z) \\ &\geq E((X - m)^2 Y) \quad [\because (X - m)^2, Z \geq 0 \quad \therefore E((X - m)^2 Z) \geq 0] \end{aligned}$$

Now $(X - m)^2 Y \geq \epsilon^2 Y$.

⟦ Because:

Y is either 0 or 1.
If $Y = 0$, then LHS $=$ RHS $= 0$.
If $Y = 1$ then $|X - m| \geq \epsilon$, and so

$$\text{LHS} = (X - m)^2 \times 1 \geq \epsilon^2 \times 1 = \text{RHS}.$$

>]]
>
> Taking expectation we have $E((X-m)^2Y) \geq \epsilon^2 E(Y)$.

এবার $E(Y)$ বার করি। যেহেতু Y খালি 0 আর 1 হতে পারে, তাই--

> Now
>
> $$E(Y) = 1 \times P(|X-m| \geq \epsilon) + 0 \times P(|X-m| < \epsilon) = P(|X-m| \geq \epsilon).$$

এবার সবকিছু মিলিয়ে পাচ্ছি--

> Hence
>
> $$\sigma^2 \geq \epsilon^2 E(Y) = \epsilon^2 P(|X-m| \geq \epsilon).$$
>
> So
>
> $$P(|X-m| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2},$$
>
> as required, since $\epsilon^2 > 0$.

■

Exercise 1: If X is a random variable with finite mean μ and finite variance σ^2, then prove that for any $k > 0$

$$P(|X-\mu| < k) \geq 1 - \frac{\sigma^2}{k^2}.$$

(2008.3d)

HINT:
এটা আসলে আগের অংকটাই, খালি দুই দিকেই complement নেওয়া হয়েছে। ■

ভালো করে Chebyshev inequality-টা লক্ষ কর--

$$\forall \epsilon > 0 \quad P(|X - E(X)| \geq \epsilon) \leq \frac{\text{var}(X)}{\epsilon^2}.$$

এখানে একজায়গায় $\geq$ এবং একজায়গায় $\leq$ আছে। যদি এদের জায়গায় $>$ আর $<$ বসাতাম, তবে কী হত? সেই নিয়েই নীচের অংকটা।

Exercise 2: ধরো X একটা random variable, যার জন্য $E(X) = \mu$ আর $\text{var}(X) = \sigma^2 < \infty$. বলে দিচ্ছি যে, নীচের তিনটে inequality-র মধ্যে কোনটা বা কোনগুলো সব সময়েই সত্য?

1. $\forall \epsilon > 0 \quad P(|X - E(X)| \geq \epsilon) < \frac{\text{var}(X)}{\epsilon^2}$
2. $\forall \epsilon > 0 \quad P(|X - E(X)| > \epsilon) \leq \frac{\text{var}(X)}{\epsilon^2}$
3. $\forall \epsilon > 0 \quad P(|X - E(X)| > \epsilon) < \frac{\text{var}(X)}{\epsilon^2}$

■

Example 2: If X is any non-negative random variable having mean m, prove that $P(X \geq m\Sigma) \leq \frac{1}{\Sigma}$ for any $\Sigma > 0$ (both discrete and continuous).[5] **(2012.3a)**

SOLUTION: অংকটায় একটা ভুল আছে। এখানে ধরে নিতে হবে যে $m > 0$. যদি $m = 0$ হয় (মানে যদি X সব সময়েই 0 হয়ে থাকে), তবে $\Sigma = 100$ নিলেই দেখবে LHS-টা হচ্ছে 1, কিন্তু RHS হয়ে যাচ্ছে $\frac{1}{100}$. সুতরাং LHS $\leq$ RHS হচ্ছে না মোটেই। কিন্তু $m > 0$ হলে সে সমস্যা হবে না।

We assume $m > 0$.

যেহেতু এখানে discrete আর continuous দুটো ক্ষেত্রেই দেখাতে বলেছে, তাই এখানেও আমরা এক ঢিলে দুই পাখি মারার জন্য Y, Z দিয়ে লিখব।

Let

$$Y = \begin{cases} 1 & \text{if } X \geq m\Sigma \\ 0 & \text{otherwise} \end{cases}$$

and $Z = 1 - Y$. Then

$$E(Y) = 1 \times P(X \geq m\Sigma) + 0 \times P(X < m\Sigma) = P(X \geq m\Sigma).$$

এবার ঠিক Chebyshev inequality-র প্রমাণের মতই এগোব, খালি variance-এর জায়গায় expectation নিয়ে--

$$\begin{aligned} m &= E(X) \\ &= E(X(Y+Z)) \quad [\because Y + Z = 1] \\ &\qquad \text{সেই পরিচিত কৌশল, } Y+Z=1 \text{ দিয়ে গুণ করে লেখা।} \\ &= E(XY) + E(XZ) \\ &\geq E(XY) \quad [\because X, Z \geq 0] \end{aligned}$$

Now $XY \geq m\Sigma Y$.

⟦ Because:

Y is either 0 or 1.
If $Y = 0$, then LHS = RHS $= 0$.
If $Y = 1$ then $X \geq m\Sigma$, and so

$$\text{LHS} = X \times 1 \geq m\Sigma \times 1 = \text{RHS}.$$

⟧

Taking expectation we have $E(XY) \geq m\Sigma E(Y)$.

So $m \geq m\Sigma E(Y) = P(X \geq m\Sigma)$.

Since $m > 0$,

$$\therefore P(X \geq m\Sigma) \leq \frac{1}{\Sigma},$$

as required.

দেখলে, $m > 0$ কোথায় লাগল? ■

26.1 কিছু প্রয়োগ

Chebyshev inequality খাটে যেকোনো random variable-এর জন্যই (variance-টা অবশ্য finite হবে)। এবার আমরা কিছু বিশেষ distribution-এর জন্য inequality-টা প্রয়োগ করে দেখব।

Example 3: A random variable X has probability density function

$$f(x) = 12x^2(1-x) \quad 0 < x < 1.$$

Compute $P(|X-m| \geq 2\sigma)$ and compare it with the limit given by Tchebycheff's inequality.[4+1] **(2011.3e)**

SOLUTION: এখানে m আর σ কী জিনিস বলে দেওয়া উচিত ছিল। যাই হোক, নিশ্চয়ই আন্দাজ করতে পেরেছ যে, আসলে m, σ^2 বলতে যথাক্রমে $E(X), \mathrm{var}(X)$ বোঝাতে চেয়েছে। প্রথমে এদের বার করে নিই।

$$m = E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 12x^3(1-x)dx = 12\int_0^1 x^{4-1}(1-x)^{2-1}dx = \cdots = \frac{3}{5}.$$

Also

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 12x^4(1-x)dx = \cdots = \frac{2}{5}.$$

So

$$\sigma^2 = \frac{2}{5} - \left(\frac{3}{5}\right)^2 = \frac{1}{25}.$$

So

$$\begin{aligned} P(|X-m| \geq 2\sigma) &= P\left(|X - \tfrac{3}{5}| \geq \tfrac{2}{5}\right) \\ &= 1 - P\left(\tfrac{1}{5} < X < 1\right) \\ &= \cdots \\ &= \frac{17}{625}. \end{aligned}$$

এবার Chebyshev inequality-র সঙ্গে তুলনা করব। প্রশ্নে "Tchebycheff" বানান দেখে খাবি খেও না যেন। ওই ভদ্রলোকের নামের সাড়ে সাঁইত্রিশরকম বানান হয়। আর একটা কথা--অংকটায় যেভাবে "limit" শব্দটা ব্যবহার করা হয়েছে, সেটা কিন্তু অংকের ব্যাকরণসম্মত নয়। বলা উচিত ছিল "bound".

By Chebyshev's inequality we have

$$P(|X-m| \geq 2\sigma) \leq \frac{\sigma^2}{(2\sigma)^2} = \frac{1}{4}.$$

এখানে আমরা Chebyshev inequality-তে $\epsilon = 2\sigma$ নিয়েছি।

Comparing we have

$$\frac{17}{625} \leq \frac{1}{4}.$$

■

Example 4: Show by Tchebycheff's inequality that in 2000 throws with a coin the probability that the number of heads lies between 900 and 1100 is at least $^{19}/_{20}$.[4] **(2004.6c)**

SOLUTION:

We assume that the coin is unbiased.
Let X be the number of heads in 2000 throws.
Then $X \sim Binomial\left(2000, \frac{1}{2}\right)$.
So $E(X) = 2000 \times \frac{1}{2} = 1000$ and $\text{var}(X) = 2000 \times \frac{1}{2} \times \left(1 - \frac{1}{2}\right) = 500$.
Hence, by Chebyshev's inequality, for any $\epsilon > 0$ we have

$$P(|X - E(X)| > \epsilon) \leq P(|X - E(X)| \geq \epsilon) \leq \frac{\text{var}(X)}{\epsilon^2}.$$

এখানে একটা সামান্য প্যাঁচ করেছি। আমাদের পৌঁছতে হবে $P(900 \leq X \leq 1100)$-এ, মানে $P(|X - 1000| \leq 100)$. তার complement হল $P(|X - 1000| < 100)$. কিন্তু Chebyshev inequality লাগাতে হলে ওই "<"-টার জায়গায় "$\leq$" চাই। সেই কারণেই আমরা $P(|X - 1000| < 100) \leq P(|X - 1000| \leq 100)$ ব্যবহার করেছি।

Putting $\epsilon = 100$

$$P(|X - 1000| > 100) \leq \frac{500}{100^2} = \frac{1}{20}.$$

So, taking complements,

$$P(900 \leq X \leq 1100) > \frac{19}{20},$$

as required.

■

Exercise 3: Use Chebyshev's inequality to show that if a die is thrown 3600 times, then the probability that the number of sixes lies between 550 and 650 is at least 4/5. [Hint: $P(|X - 600| < 50) \geq 1 - \frac{500}{50^2}$.□] **(2014.3a)**

HINT: এটাও আগের অংকটারই মত। কিন্তু এখানে উত্তরটা প্রায় বলেই দিয়েছে ওই hint-এর মধ্যে। যদি X ধরো মোট ছয় পড়ার সংখ্যা, তবে X-এর distribution হবে Binomial $\left(3600, \frac{1}{6}\right)$. এখানে ধরে নিয়েছি যে ছক্কাটার সব দিক সমান ভারী। এবার Chebyshev inequality লাগালেই উত্তর পেয়ে যাবে। ■

Example 5: If X is a $\gamma(n)$ variate, then show that

$$P(0 < X < 2n) \geq \frac{n-1}{n}.[5]$$

(2013,2007)

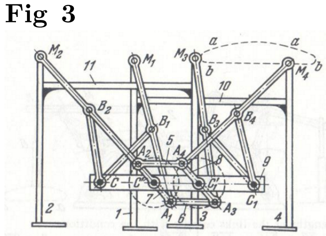
SOLUTION: Distribution বোঝানোর জন্য $\gamma(n)$ চিহ্নটা খুব প্রচলিত নয়। Gamma(n) ব্যবহার করাই ভালো।

> We know that $E(X) = var(X) = n$ using standard property of Gamma(n) distribution.
> Also from Chebyshev's inequality we know that for any $\epsilon > 0$
>
> $$P(|X - E(X)| \geq \epsilon) \leq \frac{var(X)}{\epsilon^2}.$$
>
> Taking $\epsilon = n$, we have
>
> $$P(|X - n| \geq n) \leq \frac{n}{n^2},$$
>
> or, taking complements,
>
> $$P(|X - n| < n) \geq 1 - \frac{1}{n},$$
>
> or
>
> $$P(0 < X < 2n) \geq \frac{n-1}{n},$$
>
> as required.

■

Chebyshev ভদ্রলোকের কাজকর্ম বেশ বিচিত্র। Probability নিয়ে ওনার একটা কাজ তো দেখলে, এবার আরেকটা কাজের কথা বলি, যেটা শুনলে অবাক হয়ে যাবে। জন্তুজানোয়ার, মানুষ ইত্যাদি যখন চলাফেরা করে তখন সেটা করে পা ফেলে ফেলে। কিন্তু মানুষের তৈরী কৃত্রিম যানবাহনগুলো সবই চলে চাকাজাতীয় জিনিসের সাহায্যে (বা স্লেজ গাড়ির মত ঘস্‌টে ঘস্‌টে)। কারণটা বোঝা কঠিন নয়, চাকা ঘোরানো বা স্লেজগাড়ি টানার জন্য বিরাট বুদ্ধি লাগে না, কিন্তু পা ফেলে হাঁটার জন্য বেশ কিছুটা বুদ্ধি লাগে, কোন পা কখন ফেলব, কীভাবে ভারসাম্য রাখব, ইত্যাদি। Chebyshev একটা যন্ত্রের পরিকল্পনা করেছিলেন যেটা কোনো বুদ্ধি ছাড়াই সম্পূর্ণ যান্ত্রিকভাবে চার পায় হেঁটে বেড়াতে পারে! জিনিসটা চালানো যায় দুটো হাতল ঘুরিয়ে (যেগুলোকে মোটর দিয়ে ঘোরানো যেতে পারে)। ওনার গবেষণার পাতা থেকে যন্ত্রটার একটা খসড়া তুলে দিয়েছি Fig 3-এ। ছবিটা দেখে

Fig 3



খুব কিছু বোঝা যাচ্ছে এমন নয়, কিন্তু যন্ত্রটা সত্যিই কাজ করে, এবং শুনেছি এই কায়দায় একধরণের বইয়ের আলমারী বানানো হয়েছে যেটা নাকি ঠেলা দিলে পা ফেলে ফেলে এগিয়ে যায়! Google-এ খুঁজে দেখতে পারো "Chebyshev's walking platform" দিয়ে, কিছু video-ও পাবে।

DAY 27
Limits (part 1)

আমরা এ বইয়ের গোড়াতেই statistical regularity বোঝানোর সময়ে বারবার লুডো খেলে একটা পাতার ছবি পাচ্ছিলাম, মনে আছে? যেভাবেই আমরা লুডোটা খেলছিলাম, ওই একই পাতার ছবি ফিরে ফিরে আসছিল। ম্যাজিকটার একটা সহজ সংস্করণও বলেছিলাম তখন, একটা কয়েনকে বারবার করে টস করে গেলে head পড়ার relative frequency-র গ্রাফটা $P(\text{head})$-এর দিকে এগোতে থাকে। এর প্রমাণটা এবার আমরা করব। প্রথমে পুরো গল্পটা একটু মনে করিয়ে দিই।
ধরো একটা কয়েন নিলাম, যার $P(\text{head})$ কিছু একটা, ধরো p. কয়েনটাকে বারবার টস করলাম। মনে করো, প্রথম পাঁচবারে পেলাম

head, tail, tail, head, head.

আমরা প্রতিটা টসের পর হিসেব করে দেখব head পড়ার relative frequency কত হয়েছে, মানে এ পর্যন্ত যতগুলো head পড়েছে, তাকে এ পর্যন্ত মোট টসের সংখ্যা দিয়ে ভাগ করে দেব। এইভাবে যে relative frequency-গুলো পাবে, তাদের $X_1, X_2, \ldots$ ইত্যাদি নাম দাও। যেমন, আমাদের উদাহরণে $X_1 = \frac{1}{1}$, (যেহেতু প্রথম টসেই head পড়েছে), $X_2 = \frac{1}{2}$ (যেহেতু প্রথম দুটো টসে খালি একটাই head), $X_3 = \frac{1}{3}, X_4 = \frac{2}{4} = \frac{1}{2}, X_5 = \frac{3}{5}$, ইত্যাদি।
আমরা আমাদের অভিজ্ঞতা থেকে জানি যে, X_n-গুলো ক্রমশঃই p-এর দিকে এগিয়ে যাবে। প্রথমে এই "এগিয়ে যাওয়া"-র ব্যাপারটা অংকের ভাষায় লিখতে শিখব।

27.1 Convergence in probability

ধরো, 10000 বার টস করলাম। তবে আমরা জানি যে, head পড়ার proportion নিশ্চয়ই p-এর খু-উ-উ-ব কাছে থাকবে। কিন্তু এমনও তো হতে পারে যে, সব বারই আমরা head পেয়েছি। হ্যাঁ, তার সম্ভাবনা খুবই কম, কিন্তু একেবারে অসম্ভব তো নয়! সুতরাং যেটা বলা যাচ্ছে সেটা হল $|X_{10000} - p|$ ছোটো হবার probability খুব বেশী হবে, বা ঘুরিয়ে বললে, বড় হবার probability খুব কম। সেটাকে আমরা এইভাবে লিখতে পারি--

যা খুশি একটা $\epsilon > 0$ নাও। তাহলে $P(|X_n - p| \geq \epsilon) \to 0$ হবে।

এই ব্যাপারটাকে আমরা বলি **convergence in probability**, যার বিশদ সংজ্ঞাটা চেয়েছে নীচের অংকে।

Example 6: Define the concept of convergence in probability. [1] **(2012,2008,2005,2004)**

SOLUTION:

DEFINITION: Convergence in probability

Let $\{X_n\}_n$ be a sequence of random variables and X be a random variable, all defined for the same random experiment. We say that X_n converges to X in probability, and write $X_n \xrightarrow{P} X$ if

$$\forall \epsilon > 0 \quad \lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0.$$

■

Exercise 4: একটু চিন্তা করে বুঝে নাও যে, এখানে "$\geq$"-র জায়গায় "$>$" লিখলেও সংজ্ঞাটা একই থাকত। গুছিয়ে প্রমাণ করতে পারো? না পারলে উত্তর দেখে নিও। ■

Convergence in probability-র ধারণাটা অনেকটা আমাদের সাধারণ limit-এর মতই। তাই সাধারণ limit-এর অনেক ধর্মই এখানেও খাটে। এরকম একটা উদাহরণ রয়েছে নীচের অংকটায়।

Example 7: If $X_n \xrightarrow{P} X;\quad Y_n \xrightarrow{P} Y$ as $n \to \infty$, then show that

$$(X_n \pm Y_n) \xrightarrow{P} (X \pm Y) \text{ as } n \to \infty.$$

[4] **(2012,2008,2005,2004)**

SOLUTION:

Let $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$.

Shall show $X_n \pm Y_n \xrightarrow{P} X \pm Y$, ie,

$$\forall \epsilon > 0 \quad \lim_{n\to\infty} P(|(X_n \pm Y_n) - (X \pm Y)| \geq \epsilon) = 0.$$

প্রথমেই $\forall \epsilon > 0$ আছে, তাই--

Take any $\epsilon > 0$.

Then

$$\begin{aligned}
& P(|X_n \pm Y_n - (X \pm Y)| \geq \epsilon) \\
= \; & P(|(X_n - X) \pm (Y_n - Y)| \geq \epsilon) \\
\leq \; & P(|X_n - X| + |Y_n - Y| \geq \epsilon) \quad \Big[\text{by triangle inequality}\Big]
\end{aligned}$$

এটা হল কারণ, $|(X_n - X) \pm (Y_n - Y)| \leq |X_n - X| + |Y_n - Y|$.

তাই $|(X_n - X) \pm (Y_n - Y)| \geq \epsilon$ হলে অবশ্যই $|X_n - X| + |Y_n - Y)| \geq \epsilon$-ও হবে।

$$\leq \; P(|X_n - X| \geq \epsilon/2 \text{ or } |Y_n - Y| \geq \epsilon/2)$$

কারণ দুটো জিনিসের যোগফল যদি $\geq \epsilon$ হয়, তবে তাদের একজন অন্ততঃ $\geq \frac{\epsilon}{2}$ হবেই।

নইলে, দুজনেই যদি $< \frac{\epsilon}{2}$ হত, তবে ওদের যোগফলও $< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$ হত।

$$\begin{aligned}
\leq \; & P(|X_n - X| \geq \epsilon/2) + P(|Y_n - Y| \geq \epsilon/2) \quad [\because P(A \cup B) \leq P(A) + P(B)] \\
\to \; & 0 + 0 \quad \Big[\because X_n \xrightarrow{P} X \text{ and } Y_n \xrightarrow{P} Y\Big]
\end{aligned}$$

$\because$ Any probability is ≥ 0,

$\therefore$ by the sandwich law of limit,

$$\lim_{n\to\infty} P(|X_n \pm Y_n - (X \pm Y)| \geq \epsilon) = 0,$$

as required.

এখানে স্যাণ্ডউইচ লাগানো গেল কারণ probability-টা উপর নীচ দুদিক থেকেই 0-র মধ্যে ঢুকে যাচ্ছে। ■

একইরকম যুক্তি দিয়ে নীচের অংকটাও করা যায়।

Exercise 5: If $X_n \xrightarrow{P} X; \quad Y_n \xrightarrow{P} Y$ as $n \to \infty$, then show that

$$X_n Y_n \xrightarrow{P} XY \text{ as } n \to \infty.$$

Also, if Y_n's and Y never take the value 0, then

$$\frac{X_n}{Y_n} \xrightarrow{P} \frac{X}{Y} \text{ as } n \to \infty.$$

■

27.2 Weak law of large numbers (WLLN)

আমরা আবার আমাদের মূল গল্পে ফিরে আসি। আমরা এটা দেখাতে চাইছিলাম যে, যদি একটা কয়েনকে বারবার টস করা হয়, তবে head পড়ার relative frequency ক্রমশঃ $P(\text{head})$-এর দিকে এগিয়ে যাবে। এতক্ষণ আমরা "এগিয়ে যাওয়া"-র একটা যুৎসই সংজ্ঞা খাড়া করছিলাম। সত্যিই যে এগিয়ে যাবে সেটা এবার দেখানোর পালা। এই প্রমাণটা প্রথম করেছিলেন Bernoulli নামের একজন গণিতজ্ঞ। প্রমাণটা খুবই ছোট্টো, কিন্তু পুরো ধারণাটা সৃষ্টি করতে তাঁর নাকি লেগেছিল কুড়ি বছর!

Example 8: State and prove Bernoulli's limit theorem.[5] **(2004.6ap2)**

SOLUTION:

Bernoulli's limit theorem

Let $X_1, X_2, ...$ be an IID sequence of $Bernoulli(p)$ random variables. Then
$$\frac{1}{n}\sum_1^n X_i \xrightarrow{P} p \text{ as } n \to \infty.$$

আমরা একটু আগে যে কয়েন টসে head পড়ার কথা বলছিলাম, সেটাই এখানে X_i. যদি i-th টসে head পড়ে তবে $X_i = 1$, নইলে $X_i = 0$. সুতরাং $\frac{1}{n}\sum_1^n X_i$ হল প্রথম n-খানা টসে head পড়ার relative frequency. আমরা দেখাব যে, এটা p-এর দিকে এগিয়ে যায়। প্রমাণটা শুরু করার আগে দুটো জিনিস মনে করিয়ে দিই। Bernoulli(p)-এর expectation আর variance হল যথাক্রমে p এবং $p(1-p)$.

<u>Proof:</u> Let $Y_n = \frac{1}{n}\sum_1^n X_i$.

To show $Y_n \xrightarrow{P} p$, ie,
$$\forall \epsilon > 0 \quad P(|Y_n - p| \geq \epsilon) \to 0 \text{ as } n \to \infty.$$

Take any $\epsilon > 0$.

We have $E(Y_n) = \frac{1}{n}\sum_1^n E(X_i) = \frac{1}{n}\sum_1^n p = p$.

Also
$$\begin{aligned}
\text{var}(Y_n) &= \frac{1}{n^2}\text{var}\left(\sum_1^n X_i\right) \\
&= \frac{1}{n^2}\sum_1^n \text{var}(X_i) \quad [\because X_i\text{'s independent}] \\
&= \frac{1}{n^2}\sum_1^n p(1-p) \quad [\because \text{var}(X_i) = p(1-p)] \\
&= \frac{1}{n^2} \times np(1-p) = \frac{p(1-p)}{n},
\end{aligned}$$

which is finite.

So, by Chebyshev's inequality,
$$P(|Y_n - E(Y_n)| \geq \epsilon) \leq \frac{\text{var}(Y_n)}{\epsilon^2},$$

or
$$P(|Y_n - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \to 0 \text{ as } n \to \infty,$$

as required.

■

Exercise 6: If X_n is a binomial r.v. with parameters n and p, then the sequence of r.v.'s $\{X_n/n\}$ converges in probability to p.[5] **(2013.4a)**

HINT:

ভালো করে আগের অংকের সমাধানটা পড়। সেখানে আমরা $\frac{1}{n}\sum_1^n X_i$ নিয়ে কাজ করছিলাম, যেখানে X_i-রা ছিল IID Bernoulli(p). সুতরাং $\sum_1^n X_i \sim \text{Binomial}(n,p)$. এই অংকে এই sum-টাকেই X_n বলা হয়েছে। সুতরাং এটা আসলে আগের অংকটাই। গুছিয়ে লিখে ফেলার দায়িত্ব তোমার। ■

Bernoulli's limit theorem-এ আমরা কাজ করছিলাম কয়েন টস নিয়ে। কিন্তু যে যুক্তিটা আমরা ব্যবহার করছিলাম, সেটা আরও অনেক general ক্ষেত্রেও প্রযোজ্য। আমাদের দরকার খালি একটা কিছু random variable-এর একটা sequence, যেখানে random variable-গুলো IID, এবং যাদের variance-টা finite (যাতে Chebyshev inequality লাগানো যায়)। সুতরাং কয়েন টস ভুলে গিয়ে এরকম general কোনো sequence-এর উপর একই যুক্তি লাগালে একটা জোরালো theorem পাওয়া যায়, যার পোশাকি নাম হল **weak law of large numbers (WLLN)**. নামটা শুনেই আন্দাজ করতে পারছ যে, strong law of large number বলেও সম্ভবতঃ কিছু একটা আছে, নইলে খামোখা "weak" শব্দটা আসবে কেন? সত্যিই strong law of large numbers বলে একটা জিনিস আছে। তবে সেটা আমাদের বইয়ের পাল্লার বাইরে। তাই এ বইতে আমরা কখনো সখনো খালি law of large numbers বলে ছেড়ে দিলে WLLN-ই বুঝে নিও।

Example 9: State the law of large numbers and obtain Bernoulli's theorem as a particular case of the law of large numbers for equal components.[5] **(2004.6b)**

SOLUTION:

(Weak) law of large numbers

Let $X_1, X_2, \ldots$ be an IID sequence of random variables with mean μ and finite variance. Then

$$\frac{1}{n}\sum_1^n X_i \xrightarrow{P} \mu$$

as $n \to \infty$.

অংকটায় "equal components" বলে একটা কথা ব্যবহার করা হয়েছে। ওটার উদ্দেশ্য হল IID বোঝানো।

Bernoulli's theorem as a special case: If we assume that the common distribution of the X_i's is Bernoulli(p), then $E(X_i) = (1-p)\times 0 + p \times 1 = p$.

Also $X_i^2 = X_i$, and so $E(X_i^2) = E(X_i) = p$.

Hence $\text{var}(X_i) = E(X_i^2) - \big(E(X_i)\big)^2 = p - p^2 = p(1-p)$, which is finite.

So by the (weak) law of large numbers, we have

$$\frac{1}{n}\sum_{1}^{n} X_i \xrightarrow{P} p$$

as $n \to \infty$, which is Bernoulli's theorem.

■

তার মানে লক্ষ কর যে, WLLN-এর মূল শক্তিটা আসছে Chebyshev inequality থেকেই। এই ব্যাপারটাই গুছিয়ে লিখতে বলেছে নীচের অংকটায়।

Example 10: State Chebyshev's inequality and its link with weak law of large numbers. [3] **(2014.3a)**

SOLUTION: Chebyshev inequality কী জিনিস সে তো আগেই বলেছি। এবার WLLN-এর সঙ্গে এর সম্পর্কটা বলি।

Link with WLLN: WLLN states: If $X_1, X_2, ...$ are IID random variables with $E(X_i) = \mu$ and $\text{var}(X_i) < \infty$, and

$$\overline{X}_n = \frac{1}{n}\sum_{1}^{n} X_i,$$

then $\overline{X}_n \xrightarrow{P} \mu$, ie,

$$\forall \epsilon > 0 \quad \lim_{n\to\infty} P(|\overline{X}_n - \mu| \geq \epsilon) = 0.$$

This follows from Chebyshev's inequality:

Take any $\epsilon > 0$.

Let $Y_n = \overline{X}_n$.

Then $E(Y_n) = \frac{1}{n}\sum_1^n E(X_i) = \mu$.

Let $\sigma^2 = |var(X_i) < \infty$.

Also $\text{var}(Y_n) = \frac{1}{n^2}\sum_1^n var(X_i) = \frac{\sigma^2}{n}$, since X_i's are IID

So, by Chebyshev's inequality,

$$P(|Y_n - \mu| \geq \epsilon) \leq \frac{\text{var}(Y_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0.$$

Since any probability is ≥ 0, hence WLLN follows using sandwich law of limit.

■

DAY 28
Limits (part 2)

28.1 Convergence in distribution

আমরা convergence in probability কাকে বলে শিখেছি। যদি $\{X_n\}_n$ একটা random variable-এর sequence হয়, আর X কোনো random variable হয়, তবে এর সাহায্যে আমরা বলতে পারি কখন X_n-গুলো X-এর কাছাকাছি এগিয়ে যাচ্ছে। কিন্তু এটাই এই কাজ করার একমাত্র পদ্ধতি নয়। আরেকটা পদ্ধতি আছে তাকে বলে **convergence in distribution.** সেটার সংজ্ঞা হল এইরকম--এখানেও আমরা $\{X_n\}_n$ আর X নিয়ে শুরু করব। কিন্তু এখানে random variable-গুলোকে নিয়ে সরাসরি কাজ না করে, কাজ করব ওদের CDF-দের নিয়ে। ধরো X_n-এর CDF হল F_n আর X-এর CDF হল F. তাহলে আমরা বলব যে $X_n \to X$ **in distribution** যদি $F_n(x)$-গুলো $F(x)$-এর দিকে এগিয়ে যায়। এখানে একটা ছোট্টো প্যাঁচ আছে--আমরা x-এর খালি সেইসব value নিয়েই মাথা ঘামাব যেখানে $F(x)$-টা continuous. সুতরাং সব মিলিয়ে সংজ্ঞাটা দাঁড়ালো--

> DEFINITION: **Convergence in distribution**
>
> Let $\{X_n\}_n$ be a sequence of random variables, and let X be any random variable. Let the CDF of X_n be F_n, and the CDF of Xbe F. We say that X_n converges to X in distribution if for all continuity points x of F, we have
>
> $$F_n(x) \to F(x) \text{ as } n \to \infty.$$

এই continuity-র প্যাঁচটার কেন প্রয়োজন, সেটা বোঝানো এই স্বল্প পরিসরে সম্ভব নয়। বস্তুতঃ এই সংজ্ঞাটা আমরা খালি একটা বিশেষক্ষেত্রেই প্রয়োগ করব যেখানে F-টা হবে normal distribution-এর CDF. যেহেতু সেটা সর্বত্রই continuous, তাই আলাদা করে continuity নিয়ে মাথা ঘামাতে হবে না। এই বিশেষ ক্ষেত্রটার নাম হল asymptotic normality, যার বিশদ সংজ্ঞা চেয়েছে নীচের অংকটায়।

Example 11: What is meant by asymptotically normal distribution?[1] **(2005.6b)**

SOLUTION:

> DEFINITION: **Asymptotically normal distribution**
>
> Let $\{X_n\}_n$ be a sequence of random variables. Let $\{\mu_n\}_n$ be a sequence of real numbers and let $\{\sigma_n\}_n$ be a sequence of positive numbers. Let $F_n(x)$ be the CDF of $\frac{X_n-\mu_n}{\sigma_n}$. We say that X_n is asymptotically normal with mean μ_n and standard deviation σ_n if
>
> $$\forall x \in \mathbb{R} \quad F_n(x) \to \Phi(x) \text{ as } n \to \infty.$$
>
> Here $\Phi(x)$ is the CDF of $N(0,1)$ distribution.

■

এই সংজ্ঞাটা ব্যবহার করে অংক করা মোটেই সুবিধাজনক নয়। তার অন্যতম প্রধান কারণ হল $\Phi(x)$-এর কোনো ভদ্র ফর্মুলা আমাদের জানা নেই। এই বিপদ থেকে বাঁচার একটা চমৎকার পথ আছে, যেখানে আমরা CDF নিয়ে কাজ না করে characteristic function নিয়ে কাজ করি। কায়দাটা দাঁড়িয়ে আছে Levy-Cramer continuity theorem নামের একটা theorem-এর উপর, যার মোটামুটি বক্তব্যটা এইরকম--

> আমাদের কাছে $\{X_n\}_n$ আর X আছে। X_n-এর characteristic function হল $Q_n(t)$, আর X-এর বেলায় $Q(t)$. তাহলে X_n যদি X-এ converge করে in distribution, তবে $Q_n(t) \to Q(t)$ হবে। এই হল প্রথম অংশ। দ্বিতীয় অংশটা বলছে যে, conversely, যদি $Q_n(t) \to Q(t)$ হয়, তবে X_n অবশ্যই X-এর converge করবে in distribution.

Example 12: State the limit theorem for characteristic functions. With the help of this theorem derive Poisson distribution as a limit of the binomial distribution. [Hint: The characteristic function for a binomial(n,p) variate is $(pe^{it}+1-p)^n$, and the characteristic function for a Poisson(μ) distribution is $e^{\mu(e^{it}-1)}$.□] **(2010.3e)**

SOLUTION:

> **Levy-Cramer continuity theorem**
>
> Let $\{X_n\}_n$ be any sequence of random variables and let X be any random variable. Let X_n have characteristic function $\xi_n(t)$ and X have characteristic function $\xi(t)$ for $t \in \mathbb{R}$. Then $X_n \to X$ in distribution if and only if
>
> $$\forall t \in \mathbb{R} \quad \xi_n(t) \to \xi(t)$$
>
> as $n \to \infty$.

এবার এটা প্রয়োগ করব--

> Let $X_n \sim Binomial(n,p)$.
> Then X_n has characteristic function $Q_n(t) = (pe^{it}+1-p)^n$.
> Now suppose that $np = \lambda$, a fixed number and $n \to \infty$, and so $p = \frac{\lambda}{n} \to 0$.

এবার আমরা একটা limit ব্যবহার করব, যেটা আমরা হায়ার সেকণ্ডারীতে শিখে থাকি--

$$\left(1 - \frac{a}{n}\right)^n \to e^{-a}.$$

হায়ার সেকণ্ডারীতে অবশ্য এটা শিখি $a \in \mathbb{R}$-এর জন্য। কিন্তু আসলে এটা $a \in \mathbb{C}$-এর ক্ষেত্রেও খাটে। আমরা সেটাই এখানে কাজে লাগাব।

So

$$Q_n(t) = \left(1 + \frac{\lambda(e^{it}-1)}{n}\right)^n \to e^{\lambda(e^{it}-1)},$$

which is the characteristic function of Poisson(λ) distribution.

So, by the Levy-Cramer continuity theorem, the result is proved.

■

আমরা এখানে এটা ব্যবহার করেছি--

$$\left(1 - \frac{a}{n}\right)^n \to e^{-a}.$$

এর একটা generalisation আছে এইরকম--যদি $a_n \to a$ হয় তাহলেও একই জিনিস হবে

$$\left(1 - \frac{a_n}{n}\right)^n \to e^{-a}.$$

প্রমাণটা খুবই সহজ--

$$\left(1 - \frac{a_n}{n}\right)^n = \exp\left[n \log\left(1 - \tfrac{a_n}{n}\right)\right]$$

এবার

$$n \log\left(1 - \tfrac{a_n}{n}\right) = a_n \times \frac{\log\left(1 - \frac{a_n}{n}\right)}{\frac{a_n}{n}} \to a \times -1.$$

এখানে হায়ার সেকণ্ডারীতে শেখা এই limit-টা লাগালাম--

$$\lim_{x \to 0} \frac{\log(1+x)}{x} = 1.$$

আমরা $x = -a_n/n$ নিয়েছি।
এবার যেহেতু e^x হল একটা continuous function, তাই

$$\exp\left[n \log\left(1 - \tfrac{a_n}{n}\right)\right] \to e^{-a},$$

ঠিক যেমনটা দরকার ছিল।

Example 13: If X_n is a Binomial(n, p) variate, then show that $(X_n - np)/\sqrt{npq}$, $(q = 1 - p)$ is asymptotically normal$(0, 1)$. **(2009,2007,2005)**

SOLUTION:

Let $Z = \frac{X_n - np}{\sqrt{npq}}$.

We know that X_n has characteristic function

$$Q(t) = \left(q + pe^{it}\right)^n.$$

So Z has characteristic function

$$\begin{aligned}
Q_Z(t) &= E(e^{itZ}) = E\left(\exp\left[\frac{it(X_n-np)}{\sqrt{npq}}\right]\right) = \exp\left[-\frac{itnp}{\sqrt{npq}}\right] E\left(\exp\left[\frac{itX_n}{\sqrt{npq}}\right]\right)\\
&= \exp\left[-\frac{itnp}{\sqrt{npq}}\right] Q\left(\frac{t}{\sqrt{npq}}\right)\\
&= \exp\left[-\frac{itnp}{\sqrt{npq}}\right]\left(q+p\exp\left[\frac{it}{\sqrt{npq}}\right]\right)^n\\
&= \cdots
\end{aligned}$$

এখানে কয়েকটা বিচ্ছিরি দেখতে ধাপ তোমার জন্য রেখে দিলাম। মোটামুটি ব্যাপারটা এই--
প্রথমে ওই যে, $\exp\left[-\frac{itnp}{\sqrt{npq}}\right]$-টা আছে ওটা আসলে $\left(\exp\left[-\frac{itp}{\sqrt{npq}}\right]\right)^n$.
সুতরাং পুরো জিনিসটাকে (কিছু একটা)^n আকারে লিখে নাও।
তারপর একটু কাটাকাটি করলেই পাবে--

$$\begin{aligned}
&= \left(q\exp\left[-\frac{itnp}{\sqrt{npq}}\right]+p\exp\left[\frac{itq}{\sqrt{npq}}\right]\right)^n\\
&= \cdots
\end{aligned}$$

অনেক পরিশ্রম করেছি, এবার তার ফল পাব, প্রচুর কাটাকাটি হয়ে সব কিছু সহজ হয়ে যাবে।
মনে আছে নিশ্চয়ই যে, $e^x = 1 + x + \frac{x^2}{2} +$ higher terms?
সেইটা দুটো exp-এই লাগিয়ে দাও। দেখবে constant term-টা আসছে $p+q=1$,
t-ওয়ালা term-টা কেটে যাচ্ছে, আর t^2-ওয়ালা term-টা হচ্ছে $-\frac{t^2}{2n}$. সব মিলিয়ে--

$$= \left(1+\frac{1}{n}\underbrace{\left[-\frac{t^2}{2}+\text{ higher terms}\right]}_{a_n}\right)^n,$$

where the "higher terms" $\to 0$ as $n\to\infty$.

কেন higher terms জিনিসটা 0-র দিকে যাচ্ছে, বুঝলে তো? ওদের সবারই denominator-এ n-এর একটা করে power আছে। সেই power-গুলো সবাই n-এর থেকে বেশী। তাই একটা n কমন নেওয়ার পরও খানিকটা করে n ওদের denominator-এ থেকে যাবে এবং 0-র দিকে টেনে নামাবে।

Clearly, $a_n \to -\frac{t^2}{2}$.
We know that if $a_n \to a$ then

$$\left(1+\frac{a_n}{n}\right)^n \to e^a.$$

So the characteristic function of Z converges to $e^{-t^2/2}$, which is the characteristic function of $N(0,1)$ distribution.
Hence, by Levy-Cramer continuity theorem, the result follows.

■

28.2 Central limit theorem (CLT)

এক্ষুণি যে অংকটা করলাম সেখানে আমরা $Q_Z(t)$-র চেহারাটা ব্যবহার করেছিলাম। এইটা অংকের দুনিয়ার অন্যতম আশ্চর্য একটা ব্যাপার এই যে, $Q_Z(t)$ যদি সম্পূর্ণ অন্যরকম কিছুও হত তবেও asymptotic normality অক্ষুণ্ণ থাকত। চেহারাটা একেবারেই গুরুত্বপূর্ণ কিছু নয়, খালি একটা শর্ত থাকলেই চলে। সেটা হল এই যে প্রতিটা X_n যেন n-খানা IID random variable-এর যোগফল হয়, আর সেই random variable-গুলোর variance যেন finite থাকে। ব্যস্! যেমন আমাদের অংকটায় X_n ছিল binomial, তাই ওকে কয়েকটা IID Bernoulli random variable-এর যোগফল হিসেবে ভাবা যেত। এ থেকেই আসে probabilityও statistics-এর জগতের অন্যতম সাড়াজাগানো theorem, যার নাম central limit theorem. নীচের অংকটা সে বিষয়েই।

Example 14: State the central limit theorem for the case of equal components. Show that this theorem implies the law of large numbers.[5] **(2014,2007)**

SOLUTION:

> **Central limit theorem**
>
> Let $\{X_n\}_n$ be an IID sequence of random variables with mean μ and variance $\sigma^2 < \infty$. Let
>
> $$\overline{X}_n = \frac{1}{n}\sum_1^n X_k.$$
>
> Then
>
> $$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \to N(0,1) \text{ in distribution.}$$

WLLN-এর বেলায় যেমন বলেছিলাম যে, "equal components" বলতে IID বোঝানো হচ্ছে, এখানেও তাই।
এবার আমরা CLT থেকে WLLN প্রমাণ করব। প্রমাণটা ঠিক সহজ নয়। চাইলে এইবেলা খানিকটা জল খেয়ে ছাদ থেকে ঘুরে আসতে পারো। তারপর ঠাণ্ডা মাথায় প্রমাণটা নিয়ে বোসো।

> Proof of the WLLN from the CLT:

প্রথমে লিখে নিই, কী দেখাতে হবে--

> Let X_i's be as above.
>
> To show $\overline{X}_n \xrightarrow{P} \mu$,

একটু বিশদ করে লেখা দরকার। $\overline{X}_n \xrightarrow{P} \mu$ মানে হল তুমি যাই $\epsilon > 0$ নাও না কেন $P(|\overline{X}_n - \mu| > \epsilon) \to 0$ হবে, যখন $n \to \infty$হবে। লক্ষ করো যে, এখানে probability-র ভিতরে "$\geq$" না লিখে "$>$" লিখেছি। যখন convergence in probability-র সংজ্ঞা দিয়েছিলাম, তখনই বলেছিলাম যে, এতে কিছুই বদলায় না। কিন্তু এখানে "$>$" নিয়ে কাজ করলে পরে একটু সুবিধা হবে।
কোনো কিছু 0-তে converge করার সংজ্ঞা নিশ্চয়ই মনে আছে? সেটা ছিল এইরকম--তুমি যাই $\delta > 0$ নাও না কেন n যথেষ্ট বড় হলে সেই জিনিসটার absolute value হবে $< \delta$. এই দুটো জিনিসকে এক সঙ্গে লিখে পাচ্ছি এটা--

ie,

◎ $\forall\epsilon > 0 \quad \forall\delta > 0 \quad \exists N \in \mathbb{N} \quad \forall n \geq N \quad P(|\overline{X}_n - \mu| > \epsilon) < \delta.$

তোমরা যারা আমাদের বাংলায়-বোঝানো-ইংরাজি-বই সিরিজের অন্যান্য বই পড়েছো, তারা জানো গোড়ায় ওই ◎ চিহ্নটা কেন রয়েছে। ওর মানে ওই লাইনটা প্রমাণ করা আমাদের টার্গেট। ওটার দিকে চোখ রেখেই পুরো প্রমাণটা ধাপে ধাপে তৈরী হবে। প্রথমেই আছে দুটো $\forall$, তাই প্রথম কাজ হল--

$\boxed{\forall\epsilon,\delta}$ Take any $\epsilon > 0$ and any $\delta > 0$.

আমাদের টার্গেটে এবার আছে $\exists N \in \mathbb{N}$. যখনই $\exists$ থাকে, তখনই খানিকটা রাফে অংক কষে কার্যপদ্ধতি ঠিক করে নিতে হয়। আমাদের হাতে কী আছে দেখে নিই। বলা আছে যে, $\frac{\sqrt{n}(\overline{X}_n-\mu)}{\sigma}$-টা $N(0,1)$-এ converge করে, মানে $F_n(x)$ যদি $\frac{\sqrt{n}(\overline{X}_n-\mu)}{\sigma}$-এর CDF হয়, তবে যে কোনো $x \in \mathbb{R}$-এর জন্যই $F_n(x) \to \Phi(x)$ হবে, যেখানে $\Phi(x)$ হল $N(0,1)$-এর CDF. কথা হচ্ছে $P(|\overline{X}_n - \mu| > \epsilon)$-এর সঙ্গে $F_n(x)$-এর সম্পর্ক কী? এর উত্তর হল--

$$\begin{aligned} P(|\overline{X}_n - \mu| > \epsilon) &= 1 - P(|\overline{X}_n - \mu| \leq \epsilon) \\ &= 1 - P\left(\frac{\sqrt{n}|\overline{X}_n - \mu|}{\sigma} \leq \frac{\sqrt{n}\epsilon}{\sigma}\right) \\ &= 1 - P\left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \in \left[\frac{\sqrt{n}\epsilon}{\sigma}, \frac{\sqrt{n}\epsilon}{\sigma}\right]\right) \end{aligned}$$

এইবার একটা কৌশল করব। মনে রেখো যে, U যদি একটা random variable হয়, যার CDF হল $G(u)$, তবে $P(U \in (a,b]) = G(b) - G(a)$.
তাই আমরা interval-টাকে (,]-জাতীয় করতে চাইব।

$$\leq \quad 1 - P\left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \in \left(\frac{\sqrt{n}\epsilon}{\sigma}, \frac{\sqrt{n}\epsilon}{\sigma}\right]\right)$$

এটা হল কারণ যেকোনো $a < b$-এর জন্যই $(a,b] \subseteq [a,b]$.
এইবার CDF দিয়ে লিখে ফেলা যাবে--

$$= \quad 1 - \left[F_n\left(\frac{\sqrt{n}\epsilon}{\sigma}\right) - F_n\left(-\frac{\sqrt{n}\epsilon}{\sigma}\right)\right].$$

এইটাকে শেষ পর্যন্ত $< \delta$ দেখাতে হবে। আমরা F_n-এর বিষয়ে কী জানি? খালি এইটা যে, F_n-টা Φ-তে converge করে, তার মানে n যথেষ্ট বড় হলে $F_n(x)$আর $\Phi(x)$ প্রায় সমান হবে। তাহলে উপরের জিনিসটা মোটামুটি

$$1 - \left[\Phi\left(\frac{\sqrt{n}\epsilon}{\sigma}\right) - \Phi\left(-\frac{\sqrt{n}\epsilon}{\sigma}\right)\right]$$

এইটাকে কি $< \delta$ দেখানো যায়? মনে রেখো যে $\Phi(x)$একটা CDF, সুতরাং যেকোনো CDF-এর মতই $\lim_{x\to-\infty}\Phi(x) = 0$ এবং $\lim_{x\to\infty}\Phi(x) = 1$ হবেই। সুতরাং n যথেষ্ট বড় করলে এই পদার্থটাকে 0-র যত ইচ্ছে কাছে নিয়ে যাওয়া যাবে--

Now $\exists M > 0 \quad 1 - [\Phi(M) - \Phi(-M)] < \frac{\delta}{3}$.

এখানে δ না নিয়ে $\frac{\delta}{3}$ নিয়েছি, কারণ হাজার হোক F_n তো আর Φ-এর একেবারে সমান নয়, "প্রায় সমান" মাত্র। ওই "প্রায়"-টুকুর জন্য একটু ফাঁক রেখেছি। সেই ফাঁকটা কীভাবে কাজে আসবে সেটা এক্ষুণি দেখবে।
এর পরের কাজ হল F_n আর Φ কতটা কাছাকাছি আছে তার হিসেব নেওয়া--

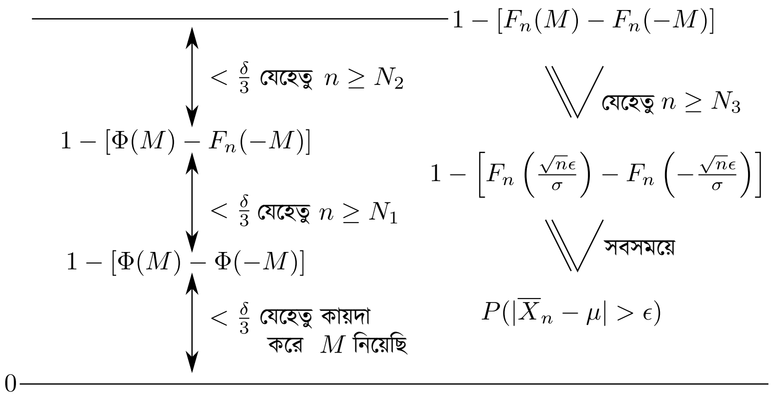


Fig 4

$$\exists N_1 \in \mathbb{N} \quad \forall n \geq N_1 \quad |F_n(M) - \Phi(M)| < \tfrac{\delta}{3}.$$
$$\exists N_2 \in \mathbb{N} \quad \forall n \geq N_2 \quad |F_n(-M) - \Phi(-M)| < \tfrac{\delta}{3}.$$

এ সবই হচ্ছে $\pm M$ বিন্দুতে। আমাদের আসল কাজ তো করতে হবে $\pm\frac{\sqrt{n}\epsilon}{\sigma}$ নিয়ে। সুতরাং এবার দেখতে হবে n কত বড় হলে $\frac{\sqrt{n}\epsilon}{\sigma} \geq M$ হয়--

$$\exists N_3 \in \mathbb{N} \quad \forall n \geq N_3 \quad \frac{\sqrt{n}\epsilon}{\sigma} \geq M.$$

এই সব কিছু মিলিয়ে আমরা $N \in \mathbb{N}$ নির্বাচন করব--

$\boxed{\exists N}$ Choose $N = \max\{N_1, N_2, N_3\} \in \mathbb{N}$.

টার্গেটটাকে আরেকবার দেখে নাও। আমরা এখন $\exists N \in \mathbb{N}$ অব্ধি এসেছি। এরপরে আছে $\forall n \geq N$, অতএব--

$\boxed{\forall n}$ Take any $n \geq N$.

ব্যস্, আর কোনো $\forall$ বা $\exists$ নেই। সুতরাং এতক্ষণ রাফে যা যা করেছি সেটা গুছিয়ে লিখে দিলেই হবে। যদি সব কিছু গুলিয়ে গিয়ে থাকে, তবে একবার Fig 4-এর উপর চোখ বুলিয়ে নিতে পারো।

Then

$$P(|\overline{X}_n - \mu| > \epsilon) = \cdots \leq 1 - \left[F_n\left(\frac{\sqrt{n}\epsilon}{\sigma}\right) - F_n\left(-\frac{\sqrt{n}\epsilon}{\sigma}\right)\right]$$

এই অব্ধি রাফেই করেছিলাম। সেখান থেকে তুলে বসিয়ে দাও।

$$\leq 1 - \big[F_n(M) - F_n(-M)\big] \quad \Big[\because n \geq N_3 \text{ and } F_n \text{ is nondecreasing}\Big]$$
$$< 1 - \big[\Phi(M) - \Phi(-M)\big] + \tfrac{\delta}{3} + \tfrac{\delta}{3} \quad [\because n \geq N_1, N_2]$$
$$< \tfrac{\delta}{3} + \tfrac{\delta}{3} + \tfrac{\delta}{3} = \delta \quad [\text{ by choice of } M],$$

as required.

■

Answers

2. কেবল দ্বিতীয়টা, কারণ $P(|X-E(X)|>\epsilon)\leq P(|X-E(X)|\geq\epsilon)$. যদি সব সময়েই $X=0$ নাও (মানে $P(X=0)=1$), তবে প্রথম আর তৃতীয়টা খাটবে না। **4.** যদি $P(|X_n-X|\geq\epsilon)\to 0$ হয়, তবে তো $P(|X_n-X|>\epsilon)\to 0$ হবেই, কারণ $0\leq P(|X_n-X|>\epsilon)\leq P(|X_n-X|\geq\epsilon)$. যদি যেকোনো $\epsilon>0$-র জন্যই $P(|X_n-X|>\epsilon)\to 0$ হয়, তবে $P(|X_n-X|>\epsilon/2)\to 0$-ও হবে। তাই $P(|X_n-X|\geq\epsilon)\to 0$ হবে।

Chapter VIII
Transformations and sampling distributions

DAY 29
Univariate transformation

আমরা জানি যে, যে কোনো random variable-এরই একটা distribution থাকে। সেই distribution-কে নানাভাবে প্রকাশ করা যায়--

- একটা কায়দা হল CDF দিয়ে। এই কায়দাটা যে কোনো random variable-এর ক্ষেত্রেই খাটে।
- আরেকটা কায়দা হল characteristic function দিয়ে। এই কায়দাটাও যে কোনো random variable-এর ক্ষেত্রেই খাটে।
- যদি random variable-টা discrete হয়, তবে PMF দিয়ে distribution-টা প্রকাশ করা যায়।
- যদি random variable-টার PDF থাকে, তবে PDF দিয়েও একই কাজ করা যায়।

এবার ধরো তোমার কাছে দুটো random variable আছে, X আর Y, যেখানে Y হল X-এর একটা function, মানে $Y = h(X)$. যদি X-এর distribution আর h দুটোই তোমার জানা থাকে তবে Y-এর distribution-ও বার করে ফেলা সম্ভব। কাজটা কঠিন হতে পারে (যেমন h যদি খুব বিচ্ছিরি কিছু হয়), কিন্তু সব সময়েই সম্ভব। একটা সহজ উদাহরণ দেখা যাক।

Example 1: ধরো X একটা discrete random variable যার PMF এইরকম--

x	-1	0	1
$P(X=x)$	0.1	0.2	0.7

তাহলে বলো তো $Y = X^2$-এর PMF কি হবে?

SOLUTION: X-এর বিভিন্ন value-র জন্য Y-এর value-গুলো বার করি--

x	-1	0	1
$y = x^2$	1	0	1

সুতরাং দেখা যাচ্ছে যে, Y খালি দুটো value নিতে পারে, 0 আর 1. এদের probability দুটো বার করতে পারলেই Y-এর distribution বেরিয়ে যাবে--

$$\begin{aligned} P(Y=0) &= P(X^2=0) = P(X=0) = 0.2 \\ P(Y=1) &= P(X^2=1) = P(X\in\{-1,1\}) = P(X=-1)+P(X=1) = 0.1+0.7 = 0.8. \end{aligned}$$

সুতরাং Y-এর PMF হল

y	0	1
$P(Y=y)$	0.2	0.8

■

এই অংকটা বেশ সহজেই হল, কারণ random variable-টা discrete ছিল। যদি continuous হত, তাহলেও করা যেত। এবার সেটাই আলোচনা করব।

Example 2: The probability density function of a random variable X is given by

$$f(x) = \begin{cases} 2xe^{-x^2} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Find the distribution function of X and density function of X^2.[5] **(2011.1e)**
SOLUTION: প্রথমে X-এর PDF-কে integrate করে X-এর CDF বার করব, তারপর সেখান থেকে Y-এর CDF বার করব। সেটাকে differentiate করে Y-এর PDF বেরোবে।

Let the distribution function of X be called $F(x)$. So $F(x) = P(X \leq x)$.
$\because X$ takes only positive values,
$\therefore \forall x < 0 \quad F(x) = 0$.
Let us take $x \geq 0$.

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= \int_{-\infty}^{x} f(t)dt \\ &= \int_{0}^{x} 2te^{-t^2}dt \\ &= \int_{0}^{x^2} e^{-u}du \quad \left[\text{putting } u = t^2\right] \\ &= (1 - e^{-x^2}). \end{aligned}$$

Thus

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ (1 - e^{-x^2}) & \text{if } x \geq 0 \end{cases}$$

এবার Y-এর CDF বার করব।

Let $Y = X^2$. Let Y have distribution function $G(y)$.
Then Y can take only nonnegative values. So $G(y) = 0$ if $y < 0$.

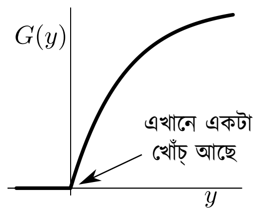


Fig 1

If $y \geq 0$, then

$$\begin{aligned} G(y) &= P(Y \leq y) \\ &= P(X \leq \sqrt{y}) \\ &= F(\sqrt{y}) \\ &= (1 - e^{-y}) \quad [\because \sqrt{y} > 0]. \end{aligned}$$

Thus

$$G(y) = \begin{cases} 0 & \text{if } y < 0 \\ (1 - e^{-y}) & \text{if } y \geq 0 \end{cases}$$

অবশেষে Y-এর PDF বার করার পালা--

Differentiating $G(y)$ we get the PDF of Y is

$$g(y) = \begin{cases} e^{-y} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

এখানে একটা সূক্ষ্ম ব্যাপার আছে। সেটা বোঝার জন্য Fig 1-এ $G(y)$-এর গ্রাফটা দ্যাখো। চোখে দেখেই বোঝা যাচ্ছে, যে $y = 0$-তে এটা একটা খোঁচ আছে, তাই ওখানে $G(y)$-টা মোটেই differentiable নয়। কিন্তু তাতে কোনো অসুবিধা নেই। যদি $G(y)$-টা countably many জায়গায় differentiable নাও হয়, তবে সেই সব জায়গায় $g(y)$ যা খুশি নেওয়া যায়, খালি ≥ 0 হলেই হল। ■

Exercise 1: The random variable X has pdf

$$f_X(x) = \frac{1}{2} e^{-|x|}, \qquad -\infty < x < \infty.$$

Find the pdf of $Y = X^2$.[4] **(2004.4b)**
HINT:
আগের অনুকরণে নিজে চেষ্টা কর। ■

Example 3: If X is uniformly distributed in the interval $(-1, 1)$ find the probability distribution of $|X|$. **(2007.2a)**
SOLUTION:

> Let the distribution function of $Y = |X|$ be $G(y)$. Clearly, Y can take values only in $(0, 1)$. So $G(y) = 0$ if $y < 0$. Also $G(y) = 1$ if $y \geq 1$.

সুতরাং বাকি রইল খালি $y \in [0, 1)$ কেসটা--

> Take any $y \in [0, 1)$.
>
> $$\begin{aligned} G(y) &= P(|X| \leq y) \\ &= P(-y \leq X \leq y) \\ &= \frac{1}{2}\int_{-y}^{y} dt \\ &= y. \end{aligned}$$

এবার সবগুলো কেস মিলিয়ে $G(y)$-টা লিখব।

> So
>
> $$G(y) = \begin{cases} 0 & \text{if } y < 0 \\ y & \text{if } 0 \leq y < 1 \\ 1 & \text{if } y \geq 1 \end{cases}$$
>
> Differentiating $G(y)$, the density of y is
>
> $$g(y) = \begin{cases} 1 & \text{if } y \in (0, 1) \\ 0 & \text{otherwise} \end{cases}$$

যেহেতু PDF-টা $(0, 1)$-এর উপরে constant, তাই আমরা বলতে পারি যে--

> Thus $|X| \sim \text{Unif}(0, 1)$.

■

Example 4: The probability density function of a random variable is symmetric around the origin. Prove that X and $-X$ have the same distribution.[6] **(2004.3b)**

SOLUTION:

> Let X have density function $f(x)$, where $f(-x) = f(x)$.
>
> Let the CDF of X be $F(x)$ and CDF of $Y = -X$ be $G(y)$.
>
> Shall show that $\forall a \in \mathbb{R} \quad F(a) = G(a)$.
>
> Take any $a \in \mathbb{R}$.

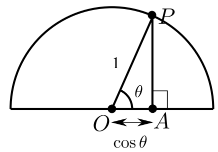


Fig 2

> Then
>
> $$\begin{aligned} G(a) &= P(-X \le a) = P(X \ge -a) \\ &= \int_{-a}^{\infty} f(x)dx \\ &= -\int_{a}^{-\infty} f(-y)dy \quad \left[\text{putting } y = -x\right] \\ &= \int_{-\infty}^{a} f(-y)dy \\ &= \int_{-\infty}^{a} f(y)dy \quad [\because f(-y) = f(y)] \\ &= P(X \le a) = F(a). \end{aligned}$$
>
> Thus X and $-X$ have the same CDF, and hence the same distribution, as required.

■

Example 5: A point is chosen at random on a semi-circle having centre at origin and radius unity and projected on the diameter. Prove that the distance of the point of projection from the centre has the probability density function

$$\frac{1}{\pi\sqrt{1-x^2}}, \quad -1 < x < 1.$$

[5] **(2008.2c)**
SOLUTION: এই অংকটায় ছবি আঁকা অপরিহার্য। Fig 2 দ্যাখো। একটা semicircle রয়েছে, যার centre-টার নাম দিয়েছি O. এবার semicircle-টার উপর random-ভাবে একটা point নেওয়া হয়েছে P. সেখান থেকে diameter-এর উপর একটা লম্ব টানা হয়েছে A-তে। OA দূরত্বটাকে নাম দিলাম X (এখানে দূরত্বটার সঙ্গে একটা চিহ্ন জড়িত আছে, A যদি O-র ডানদিকে থাকে, যেমন ছবিতে আছে, তবে দূরত্বটা positive নেওয়া হবে, নইলে negative)। যেহেতু P ছিল random, তাই X একটা random variable. এটার PDF বার করা নিয়েই অংকটা।

> Let the situation be as shown in Fig 2. Here $\theta \sim \text{Unif}(0,\pi)$.

এইটা পেলাম, কারণ P-কে semicircle-টার উপর random-ভাবে নেওয়া হয়েছে। এখানে OPA হল একটা right angled triangle, তাই $X = \cos\theta$, যেহেতু semicircle-এর radius বলা আছে 1.

> We want to find the density of $X = \cos\theta$.
> Let the CDF of X be $F(x)$.
> Clearly X must lie between -1 and 1.
> So $F(x) = 0$ if $x < -1$, and $F(x) = 1$if $x \geq 1$.

সুতরাং মাথা ঘামাতে হবে $x \in [-1, 1)$ নিয়ে--

> Take any $x \in [-1, 1)$.
>
> $$\begin{aligned} F(x) &= P(\cos\theta \leq x) \\ &= P(\theta \geq \cos^{-1} x) \quad \Big[\because \cos\theta \text{ is decreasing for } \theta \in [0, \pi] \Big] \end{aligned}$$
>
> কেন "$\leq$"-টা উল্টে "$\geq$" হয়ে গেল সেটা Fig 2 দেখে ভালো করে বুঝে নাও।
>
> $$\begin{aligned} &= P(\theta \in [\cos^{-1} x, \pi]) \\ &= \frac{\pi - \cos^{-1} x}{\pi} \quad [\because \theta \sim \text{Unif}(0, \pi)] \\ &= 1 - \frac{\cos^{-1} x}{\pi}. \end{aligned}$$
>
> So
>
> $$F(x) = \begin{cases} 0 & \text{if } x < -1 \\ 1 - \frac{\cos^{-1} x}{\pi} & \text{if } -1 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$
>
> Differentiating $F(x)$, the required density of X is
>
> $$f(x) = \begin{cases} \frac{1}{\pi\sqrt{1-x^2}} & \text{if } -1 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$
>
> as required.

■

এ কয়টা অংকে বারবারই আমরা একই কায়দা লাগাচ্ছিলাম। প্রতি ক্ষেত্রেই আমাদের একটা করে random variable-এর distribution দেওয়া থাকছিল, আর তার কোনো একটা function-র distribution বার করতে হচ্ছিল। ধরো প্রথম random variable-টা X, আর তার function-টা Y. তাহলে Y-এর PDF বার করার জন্য আমরা প্রথমে CDF-টা বার করে নিচ্ছিলাম এইভাবে--$P(Y \leq y)$-কে $P(X \in$ কোনো set$)$ আকারে লিখে নিচ্ছিলাম। তারপর এটা integrate করে বার করছিলাম। এইভাবে Y-এর CDF পেয়ে গেলে সেটাকে differentiate করে Y-এর PDF বার করা যাচ্ছিল। এই কাজটাই আরও একবার করব নীচের অংকে।

Example 6: Let X be a continuous random variable having its probability density function (pdf) $f_X(x)$. Let $y = g(x)$ be a continuously differentiable function of x. Show that the pdf $f_Y(y)$ of $Y = g(X)$ is given by

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|.$$

[7] **(2004.4a)**

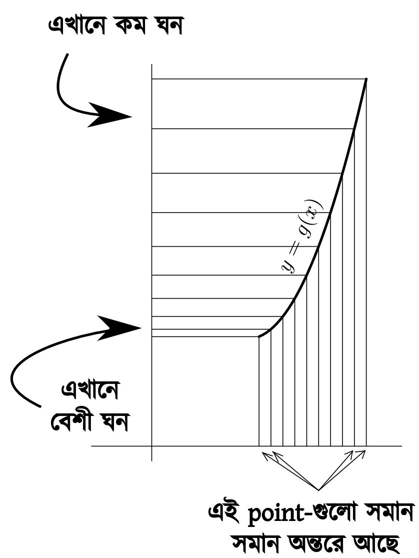
SOLUTION: অংকটা ভুল। এখানে g-কে one-one হতে হবে, এবং g^{-1}-কে differentiable হতে হবে।

> We are given $Y = g(X)$.
>
> We shall assume that g is a one-one function, and that g^{-1} is differentiable.
>
> Since g is continuous, hence this means g is either strictly increasing or strictly decreasing.

দুটো কেসে ভেঙে এগোই--

> Case 1: Let g be strictly increasing.
>
> Let CDF of X be $F_X(x)$ and the CDF of Y be $F_Y(y)$.
>
> Then
>
> $$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(g(X) \le y) \\ &= P(X \le g^{-1}(y)) \quad \left[\because g \text{ is strictly increasing} \right] \\ &= F_X(g^{-1}(y)). \end{aligned}$$
>
> Now, g^{-1} is again a differentiable function, and so differentiating F_Y

Fig 3



> we get density of Y as
>
> $$f_Y(y) = F'_Y(y) = F'_X(g^{-1}(y))\frac{dg^{-1}(y)}{y} = f_X(y)\frac{dx}{dy}.$$

অন্য কেসটাও একইরকম--

> Case 2: Let g be strictly decreasing.
>
> Similarly we get here
>
> $$f_Y(y) = -f_X(y)\frac{dx}{dy}.$$
>
> Thus, in both the cases,
>
> $$f_Y(y) = f_X(y)|[|\frac{dx}{dy}|]|,$$
>
> as required.

■

এই যে অংকটা করলাম সেটাকে ছবি দিয়ে সুন্দর করে ভাবা যায়। কায়দাটা পরেও কাজে লাগবে, তাই বলে রাখি। Fig 3 দ্যাখো। এখানে আমরা $y = g(x)$-এর গ্রাফ দেখিয়েছি। প্রথমে x-axis বরাবর আমরা কিছু point নিয়েছিলাম, সমান সমান অন্তরে। লক্ষ কর যে, g লাগানোর ফলে y-axis বরাবর point-গুলো কিন্তু আর সমান সমান অন্তরে নেই, কোথাও বেশী ঘন কোথাও কম ঘন হয়ে আছে। এই ঘনত্বের পরিমাণটা নির্ভর করছে $|g'(x)|$-এর উপরে। যেখানে $|g'(x)|$ বেশী (মানে গ্রাফটা বেশী খাড়া) সেখানে point-গুলো বেশী ছড়িয়ে গেছে। আবার যেখানে $|g'(x)|$-টা কম সেখানে point-গুলো আরো কাছাকাছি চলে এসেছে। সেই কারণে Y-এর PDF-টা $|g'(x)|$-এর উপরে নির্ভরশীল হয়ে পড়ছে। যদি $|g'(x)|$ কম হয় তবে PDF-টা বেড়ে যাচ্ছে, এবং $|g'(x)|$ বাড়লে PDF-টা কমছে। সেই জন্যেই অংকটায় $f_Y(y)$-এর মধ্যে $|[|\frac{dx}{dy}|]|$ এসেছিল, কারণ $\frac{dx}{dy} = 1/g'(x)$.
এই যে জিনিসটা শিখলাম এটা একটা গুরুত্বপূর্ণ theorem--

Jacobian formula (univariate)

Let X be a random variable with density $f_X(x)$. Let g be a one-one function with differentiable inverse. Then the density of $Y = g(X)$ is

$$f_Y(y) = f_X(g^{-1}(y))|[|\frac{dg^{-1}(y)}{dy}|]|.$$

এটা দেখতে হয়তো বেশ বিচ্ছিরি লাগছে, কিন্তু একটু তলিয়ে দেখলেই বুঝবে যে ব্যাপারটা তেমন জটিল কিছু নয়। ওই $g^{-1}(y)$-টা আসলে x ছাড়া আর কিছুই নয়। সুতরাং ডানদিকে যেটা লেখা হয়েছে সেটা আসলে হল $f_X(x)|[|\frac{dx}{dy}|]|$. ব্যস্। কিন্তু আমরা যখন Y-এর density বার করব, তখন সবকিছু y দিয়ে লিখতে হবে। সেই জন্যেই x-এর জায়গায় $g^{-1}(y)$ লিখতে হয়েছে।

এই theorem-টার একটা প্রয়োগ দেখি এবার।

Example 7: A random variable X is uniformly distributed over the interval $(0,2)$. Find the distribution function of the larger root of the quadratic equation $t^2+2t-X=0$.[5] **(2011.2b)**

SOLUTION:

Here X has PDF

$$f_X(x)=\begin{cases}\frac{1}{2} & \text{if } x\in(0,2)\\ 0 & \text{otherwise.}\end{cases}$$

The roots of the quadratic are

$$\frac{-2\pm\sqrt{4+4X}}{2}=-1\pm\sqrt{1+X},$$

which are real because $X>0$. The larger root is

$$-1+\sqrt{1+X}=g(X),\ \text{say.}$$

Here $g(x)$ is strictly increasing with inverse $g^{-1}(y)=(y+1)^2-1$, which is differentiable with derivative $2(y+1)$.

এই $g^{-1}(y)$-টা কী করে পেলাম, বুঝে নাও। আমরা প্রথমে লিখে নিলাম $y=g(x)=-1+\sqrt{1+x}$. এবার x-কে y দিয়ে লিখলাম, $x=(y+1)^2-1$. এটাই হল $g^{-1}(y)$.

So, by the Jacobian formula, $Y=g(X)$ has density given by

$$f_Y(y)=f_X(g^{-1}(y))|[|\frac{dg^{-1}(y)}{dy}|]|=\begin{cases}\frac{1}{2}\times|2(y+1)| & \text{if } (y+1)^2-1\in(0,2)\\ 0 & \text{otherwise}\end{cases}$$

এইটা পেয়েছি চোখকান বুঁজে $f_X(x)$-এর ফর্মুলায় x-এর জায়গায় সর্বত্র $g^{-1}(y)$, অর্থাৎ $(y+1)^2-1$ লিখে। আর সঙ্গে ওই $|[|\frac{d}{dy}g^{-1}(y)|]|$-টা তো আছেই।

ie,

$$f_Y(y)=\begin{cases}y+1 & \text{if } y\in(0,\sqrt{3}-1)\\ 0 & \text{otherwise}\end{cases}$$

■

Exercise 2: If a random variable X is a normal(m,σ) variate where m is the mean and σ is the standard deviation, then show that $aX+b$ is also a normal variate. Find its parameters.[2+3] **(2014.2c)**

HINT:

এই অংকটা Jacobian formula-র সহজ একটা প্রয়োগ। পুরোটা করে দেব না। খালি বলে রাখি যে, অংকটায় একটা ভুল আছে, এখানে ধরে নিতে হবে যে, $a\neq 0$. সেটা বোঝাই যাচ্ছে, কারণ $a=0$ হলে সব সময়েই $Y=b$ হবে।

অংকটা অবশ্য Jacobian formula না ব্যবহার করে characteristic function দিয়েও করা যায়। ■

এবারের অংকে দুই ধরণের Beta distribution লাগবে। এদের PDF-গুলো মনে করিয়ে দিই। Type 1 beta(m, n)-এর PDF হল

$$f(x) = \begin{cases} \frac{1}{B(m,n)} x^{m-1}(1-x)^{n-1} & \text{if } x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

আর Type 2 beta(m, n)-এর বেলায়

$$f(x) = \begin{cases} \frac{1}{B(m,n)} \frac{x^{m-1}}{1+x^{m+n}} & \text{if } x \in (0,\infty) \\ 0 & \text{otherwise} \end{cases} .$$

Example 8: If X is $\beta_1(m, n)$ variate, show that $\frac{X}{1-X}$ is $\beta_2(m, n)$ variate.[5] **(2013.2c)**

SOLUTION:

Let $g(x) = \frac{x}{1-x}$ and $Y = g(X)$.
We know that the density of X is

$$f_X(x) = \begin{cases} \frac{1}{B(m,n)} x^{m-1}(1-x)^{n-1} & \text{if } x \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

Now $g(x)$ is a one-one function and $x = g^{-1}(y) = \frac{y}{1+y}$.
So, by the Jacobian formula, the density of Y is

$$f_y(y) = f_x(T^{-1}(y))|[|\frac{d}{dy}g^{-1}(y)|]|.$$

Also

$$\frac{d}{dy}g^{-1}(y) = \frac{1}{(1+y)^2}.$$

So

$$f_y(y) = \begin{cases} \frac{1}{B(m,n)} \left(\frac{y}{1+y}\right)^{m-1} \left(\frac{1}{1+y}\right)^{n-1} \frac{1}{(1+y)^2} & \text{if } \frac{y}{1+y} \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

এবার এটাকে সুন্দর করে লিখব। যাদের x, y নিয়ে বেশী নাড়াচাড়া করতে হাত কাঁপে, তারা এই গুছিয়ে লেখাটা রাফে কয়েকধাপে ভেঙে করলে সুবিধা হবে। আমি এক ধাপেই লিখে দিলাম--

or

$$f_Y(y) = \begin{cases} \frac{1}{B(m,n)} \frac{y^{m-1}}{1+y^{m+n}} & \text{if } y \in (0,\infty) \\ 0 & \text{otherwise} \end{cases} ,$$

which is the density of $\beta_2(m, n)$.
Hence the result.

■

Exercise 3: If X is a $\beta_2(\ell, m)$ variate, then show that $Y = \frac{1}{X}$ is a $\beta_2(m, \ell)$ variate.[5] **(2012.2e)**

HINT: আগের অংকের অনুকরণে করো। ■

DAY 30
Bivariate transformation

এতক্ষণ আমরা Univariate transformation আলোচনা করেছি, অর্থাৎ শুরু করছিলাম একটা X নিয়ে, সেটা থেকে একটা নতুন Y পাচ্ছিলাম। চাইলে আমরা দুটো random variable নিয়েও শুরু করতে পারতাম, ধরো X, Y. তারপর তাদের মিলিয়ে-মিশিয়ে দুটো নতুন random variable বার করতে পারতাম, U, V. যেমন $U = X + Y$ আর $V = XY$, এইরকম। সেখানেও একইরকম প্রশ্ন করা চলত--X, Y-এর joint distribution জানা থাকলে কী করে U, V-র distribution বার করা যাবে? এখানেও Jacobian-এর কায়দাটা খাটবে।

Jacobian formula (bivariate)

Let X, Y be two random variables with joint PDF $f_{XY}(x, y)$. Let $(U, V) = g(X, Y)$, where $g(x, y)$ is a one-one function, and g^{-1} is differentiable. Then (U, V) has joint PDF given by

$$f_{UV}(u, v) = f_{XY}(g^{-1}(u, v))|[|det\left(\frac{\partial(x, y)}{\partial(u, v)}\right)|]|.$$

এটার মানে কী ধাপে ধাপে বোঝা যাক--

1. দুটো random variable আছে X, Y তাদের থেকে আরও নতুন দুটো random variable বানানো হয়েছে U, V. ধরো, $U = X + Y$ আর $V = X - Y$. এটাকেই সংক্ষেপে লেখা হয়েছে $(U, V) = g(X, Y)$. যেমন আমাদের ক্ষেত্রে $g(x, y) = (x + y, x - y)$.

2. এই g-টাকে one-one হতে হবে, অর্থাৎ কিনা U আর V জানা থাকলে যেন X আর Y বার করে ফেলা যায়। সেইটা আমাদের ক্ষেত্রে পরীক্ষা করা যাক। $u = x + y$ আর $v = x - y$ জানা থাকলে $x = (u + v)/2$ এবং $y = (u - v)/2$ হবে। এই যে উলটো function-টা পেলাম একেই বলেছে $g^{-1}(u, v)$. মানে আমাদের ক্ষেত্রে $g^{-1}(u, v) = \left(\frac{u+v}{2}, \frac{u-v}{2}\right)$.

3. Theorem-টায় ধরে নেওয়া হয়েছে যে g^{-1}-টা differentiable. এইবার বার করতে হবে এই $g^{-1}(u, v)$-এর derivative. যেহেতু এখানে একাধিক variable থেকে একাধিক variable পাওয়া যাচ্ছে (u, v থেকে x, y), তাই derivative-এর একটা বাহারি নাম দেওয়া হয়--**Jacobian.** লেখার সময়ে ওটাকেই লিখেছি $\frac{\partial(x,y)}{\partial(u,v)}$. এটা একটা matrix--

$$\begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}.$$

যেমন, আমাদের উদাহরণে Jacobian-টা হবে

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

4. এই matrix-টার determinant বার করে, তার absolute value নিতে হবে। সেটাই হল
$$||det\left(\frac{\partial(x,y)}{\partial(u,v)}\right)||.$$
আমাদের উদাহরণে determinant-টা হবে $-\frac{1}{2}$. তার absolute value নিয়ে পাব $\frac{1}{2}$.

5. এই বার সব কিছু মিলিয়ে বার করতে হবে
$$f_{UV}(u,v) = f_{XY}(g^{-1}(u,v))||det\left(\frac{\partial(x,y)}{\partial(u,v)}\right)||,$$
যেটা আমাদের প্রয়োজনীয় density-টা।

মাথা নিশ্চয়ই বোঁ বোঁ করে ঘুরছে? এই বিদঘুটে ধাপগুলো কেন কাজ করে, সেটা নিয়ে পরে মাথা ঘামাব। আপাততঃ একটা অংক কষে ধাপগুলো বুঝে নিই।

Example 9: If X and Y are independent variates both uniformly distributed over $(0,1)$, find the distributions of $X+Y$ and $X-Y$.[5] **(2008.3a)**

SOLUTION: এখানে প্রশ্নের ভাষায় একটা ভুল আছে। বলেছে "find the distributions of $X+Y$ and $X-Y$"। কিন্তু আসলে বলতে চেয়েছিল "find the joint distribution of $X+Y$ and $X-Y$."
প্রথমে (X,Y)-এর joint PDF-টা কী দেখে নিই--

> X has PDF $f_X(x) = \begin{cases} 1 & \text{if } x\in(0,1) \\ 0 & \text{otherwise} \end{cases}$.
>
> Y has PDF $f_Y(y) = \begin{cases} 1 & \text{if } y\in(0,1) \\ 0 & \text{otherwise} \end{cases}$.
>
> Since X, Y are independent, hence the joint PDF of (X,Y) is
> $$f_{XY}(x,y) = f_X(x)f_Y(y) = \begin{cases} 1 & \text{if } (x,y)\in S \\ 0 & \text{otherwise} \end{cases},$$
> where $S = \{(x,y) \ : \ 0 < x, y < 1\}$.

এবার U,V কী করে পাওয়া যাচ্ছে দেখি--

> Here $(U,V) = g(X,Y)$, where $g(x,y) = (x+y, x-y)$.
> This g is one-one with g^{-1} given by
> $$\begin{aligned} X &= \tfrac{1}{2}(U+V), \\ Y &= \tfrac{1}{2}(U-V). \end{aligned}$$
> This g^{-1} is differentiable, and its Jacobian is
> $$\frac{\partial(x,y)}{\partial(u,v)} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

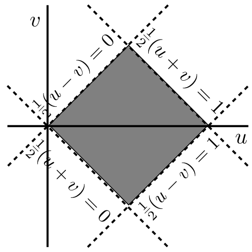


Fig 4

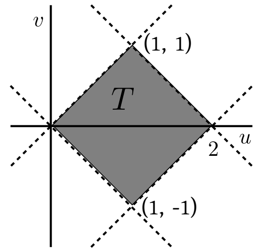


Fig 5

So

$$||[|det\left(\frac{\partial(x,y)}{\partial(u,v)}\right)|]| = \frac{1}{2}.$$

So, using the Jacobian formula, the PDF of (U,V) is

$$f_{UV}(u,v) = f_{XY}(g^{-1}(u,v))|[|det\left(\frac{\partial(x,y)}{\partial(u,v)}\right)|]| = \begin{cases} \frac{1}{2} & \text{if } \left(\frac{u+v}{2}, \frac{u+v}{2}\right) \in S \\ 0 & \text{otherwise} \end{cases}$$

$0 < \frac{1}{2}(U+V) < 1$ and $0 < \frac{1}{2}(U-V) < 1.$

Solving these inequalities graphically as in Fig 4 we get $(u,v) \in T$, where

$$T = \{(u,v) \;:\; 0 < u+v < 2,\; 0 < u-v < 2\}.$$

T is shown in Fig 5. So the required PDF is

$$f_{UV}(u,v) = \begin{cases} \frac{1}{2} & \text{if } (u,v) \in T \\ 0 & \text{otherwise} \end{cases}$$

■

এবার সংক্ষেপে আলোচনা করা যাক যে, Jacobian ফর্মুলাটা কেন কাজ করে। আমরা (x,y)-এর উপর g লাগিয়ে (u,v) পেয়েছিলাম। এটাকে এইভাবে ভাবো--

একটা গ্রাফ কাগজ আছে, সেটাকে দোমড়ানো হয়েছে। দোমড়ানোর আগে ছিল Fig 6-এর মত, দোমড়ানোর পরে Fig 7-এর মত। কীভাবে দোমড়ানো হচ্ছে সেটা ঠিক করে দিচ্ছে g. অর্থাৎ যদি $(u,v) = g(x,y)$ হয়, তবে দোমড়ানোর আগে যে বিন্দুটা ছিল (x,y)-তে সেটা দুমড়ে গিয়ে উপস্থিত হবে (u,v)-তে। এই দোমড়ানোটা নানারকমের হতে পারে, আমাদের ছবিতে আমরা একটা উদাহরণ দেখিয়েছি। লক্ষ কর যে দোমড়ানোর ফলে কোনো জায়গায় ছকগুলো চিমসে গেছে, কোথাও বা

Fig 6

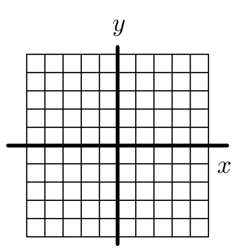
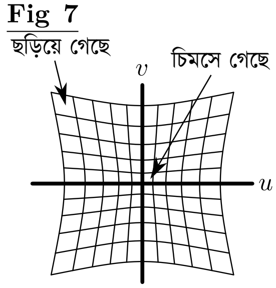


Fig 7



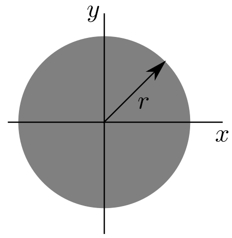


Fig 8

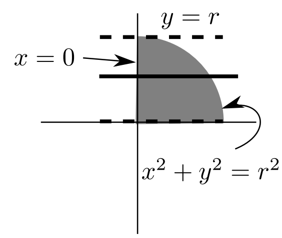


Fig 9

বিস্তৃত হয়ে গেছে (ব্যাপারটা কল্পনা করার সময়ে কাগজ না ভেবে রবারের চাদর বলে ভাবলে সুবিধা হবে)। যেখানে বিস্তৃত হয়ে গেছে, সেখানে density কমে যাবে (কারণ একই probability বেশী জায়গায় ছড়িয়ে যাচ্ছে), আবার যেখানে চিমসে গেছে সেখানে density বেড়ে যাবে। তার মানে উত্তরটা হবে

(x,y)-এর density $\times$ ওইখানে ছকটা কতটা চিমসে গেছে।

ওই "চিমসে যাওয়া"-র পরিমাণটাই বার করার কৌশল হল Jacobian-এর determinant-এর absolute value. ব্যাপারটা এইরকম--দোমড়ানো গ্রাফ কাগজের উপর একটা বিন্দু নাও। সেখানে একটা ছোটো square আঁকো। দোমড়ানোর আগে হয়তো সেটা square ছিল না মোটেই। যদি দোমড়ানোর আগে আর পরে area দুটো হয় Δ_1 আর Δ_2, তবে Jacobian-এর determinant-এর absolute value-টা হবে Δ_1/Δ_2. এই আশ্চর্য ব্যাপারটা কেন হয়, সেটা জানতে হলে তোমাকে calculus of several variables ভালো করে বুঝতে হবে। সে আলোচনা এই বইয়ের পাল্লার বাইরে।
অবশ্য যদি দুটো random variable-এর function হিসেবে খালি একটা random variable দেয়, তবে আগের মত CDF-এর পদ্ধতিতে এগোনোই সহজ হতে পারে। যেমন নীচের অংকটায়।

Example 10: X_1 and X_2 variables are two independent random variables each having the same probability density function

$$2xe^{-x^2}, \quad 0 < x < \infty.$$

Find the probability density function of $\sqrt{X_1^2 + X_2^2}$.[6] **(2013.3c)**

SOLUTION:

Since X, Y are independent, hence their joint PDF is

$$f_{XY}(x,y) = \begin{cases} 2xe^{-x^2} \times 2ye^{-y^2} = 4xye^{-x^2-y^2} & \text{if } x,y>0 \\ 0 & \text{otherwise} \end{cases}$$

Let $R = \sqrt{X_1^2 + X_2^2}$.
Shall find the CDF $F(r) = P(R \leq r)$.
Clearly, $R > 0$. So $F(r) = 0$ if $r < 0$.
Take any $r \geq 0$.
Then

$$F(r) = P(R \leq r) = P((X,Y) \in S),$$

where S is the disc with radius r centred at origin as shown in Fig 8.

This is

$$\iint_S f_{XY}(x,y)dxdy = \int_0^r \int_0^{\sqrt{r^2-y^2}} 4xye^{-x^2-y^2}dxdy$$

লক্ষ কর যে, S-এর যে অংশটা প্রথম quadrant-এ আছে, আমরা খালি সেটুকু নিয়েই মাথা ঘামাব। এই অংশটার উপর ষষ্ঠ অধ্যায়ের স্যাণ্ডউইচ-শিককাবাব কায়দা লাগিয়েছি। Fig 9 দ্যাখো।

$$= 4\int_0^r \left[ye^{-y^2} \int_0^{\sqrt{r^2-y^2}} xe^{-x^2}dx \right] dy$$

$$= \cdots$$

এইখানে $u = -x^2$ নিলে সুবিধা হবে।

$$= 2\int_0^r y\left[e^{-y^2} - e^{-r^2}\right] dy$$

$$= \cdots$$

এবার $v = -y^2$ নাও।

$$= 1-(r^2+1)e^{-r^2}$$

So the CDF of R is

$$F(r) = \begin{cases} 0 & \text{if } r \leq 0 \\ 1-(r^2+1)e^{-r^2} & \text{otherwise} \end{cases}$$

এবার এটাকে differentiate করলেই PDF পেয়ে যাব--

Differentiating $F(r)$, we get the required density as

$$f(r) = \cdots = \begin{cases} 2r^3e^{-r^2} & \text{if } r > 0 \\ 0 & \text{otherwise} \end{cases}$$

■

আমরা যেমন Jacobian formula-র univariate সংস্করণকে generalise করে bivariate সংস্করণ পেয়েছিলাম, তেমনি আরো generalise করলে একটা multivariate সংস্করণও পাওয়া যায়--

Jacobian formula (multivariate)

Let $\vec{X} \equiv (X_1,...,X_p)$ be a p-dimensional random variable with joint PDF $f_{\vec{X}}(x_1,..,x_n)$. Let $\vec{Y} \equiv (Y_1,...,Y_p) = g(\vec{X})$, where $g : \mathbb{R}^p \to \mathbb{R}^p$ is a one-one function, and g^{-1} is differentiable. Then $\vec{Y}$ has joint PDF given by

$$f_{\vec{Y}}(y_1,...,y_p) = f_{\vec{X}}(g^{-1}(y_1,...,y_p))|[|det\left(\frac{\partial(x_1,...,x_p)}{\partial(y_1,...,y_p)}\right)|]|.$$

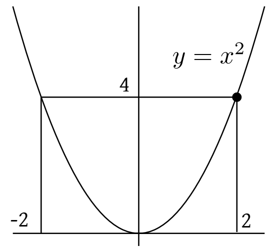


Fig 10

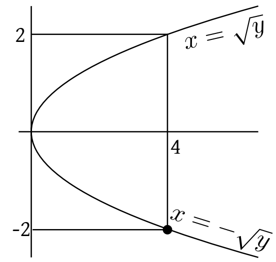


Fig 11

এটা দেখতে খুব যে নয়নাভিরাম, এমন বলছি না। কিন্তু bivariate সংস্করণটা বুঝে থাকলে এটা খুব অদ্ভুত লাগার কথা নয়। এখানে $\frac{\partial(x_1,\ldots,x_p)}{\partial(y_1,\ldots,y_p)}$ হল একটা $p\times p$ সাইজের matrix, যার (i,j)-th entry হল $\frac{\partial x_i}{\partial y_j}$.
Jacobian formula-র এই রূপটা এই অধ্যায়ে পরে খালি একবার উল্লেখ করব। তাছাড়া এই বইতে একে নিয়ে আর মাথা ঘামাব না।

30.1 যদি one-one না হয়?

আমরা তিন রকমের Jacobian ফর্মুলা শিখলাম। প্রথম সংস্করণটা কাজে লাগে যখন একটা random variable থেকে একটা random variable পাওয়া যায়। আর দ্বিতীয়টা কাজে লাগে যখন দুটো থেকে দুটো পাওয়া যায়। তৃতীয়টা সবচেয়ে বেশী general, যখন p-খানা থেকে p-খানা variable পাওয়া যায়। কিন্তু সব ক্ষেত্রেই একটা শর্ত লাগে, সেটা হল one-one হতে হবে, এবং inverse-টাকে differentiable হতে হবে। যদি one-one না হয়, তবেও কিন্তু অনেক সময়েই অংকটা একই কায়দায় করা যায়, খালি একটা বাড়তি কৌশল লাগে। সেটা বলার আগে one-one শর্তটা কোথায় লাগছে, সেটা একটা উদাহরণ দিয়ে বোঝা যাক।

Example 11: ধরো $Y=g(x)=X^3$. তুমি X-এর density জানো। বার করতে চাই Y-এর density, ধরো $y=8$-এ। তার জন্য এইভাবে এগোব। প্রথমে দেখব x-এর কোনো value-তে $y=8$ হয়। উত্তর হল $x=2$-তে। এই যে x-এর ঠিক একটাই value পেলাম এখানেই one-one কাজে লাগল। এবার তোমার উত্তর পাবে সেই $x=2$-তে X-এর density-কে Jacobian দিয়ে গুণ করলে। ■

ঠিক একইভাবে এগোবার চেষ্টা করা যাক, যদি one-one না হয়--

Example 12: এবার ধরো $Y=g(X)=X^2$. আর X-এর density ধরো $f_X(x)$. তোমাকে বার করতে হবে $f_Y(4)$.

SOLUTION: প্রথমে দ্যাখো x-এর কোন কোন value-র জন্য $y=4$ হয়। উত্তর হল $x=-2$ আর $x=2$. Fig 10 দ্যাখো। এইটা হল $g(x)$-এর গ্রাফ। যদি এটাকেই উল্টে ভাবো (মানে x- আর y-axis-এর স্থানবিনিময় করে) তবে পাবে g^{-1}-এর গ্রাফ (Fig 11)। দুঃখের বিষয়, g-টা one-one না হওয়ায় g^{-1} আসলে একটা function-ই নয়, কারণ একই y-এর value-তে একাধিক x-এর value সম্ভব। যেমন আমাদের $y=4$-এর জন্য $x=2$ আর $x=-2$ দুইই হতে পারে। এই দুটো জায়গাতেই আমরা আলাদা করে $\frac{dx}{dy}$ বার করব। খালি $x=-2$-তে ব্যবহার করব $x=-\sqrt{y}$, আর $x=2$-তে $x=\sqrt{y}$. সুতরাং $x=-2$ থেকে আসবে $f_X(-2)|[|-\frac{1}{2\sqrt{4}}|]|$, আর $x=2$-তে $f_X(2)|[|\frac{1}{2\sqrt{4}}|]|$. উত্তরটা পাওয়া যাবে এদের দুজনকে যোগ করে দিলে--

$$f_Y(4)=f_X(-2)\times\frac{1}{2}+f_X(2)\times\frac{1}{2}.$$

কেন এই কৌশলটা কাজ করে সেই প্রসঙ্গে আর যাব না। ■

DAY 31
Sampling distribution

31.1 কী জিনিস?

আমরা দেখলাম যে, দুটো random variable যদি দেওয়া থাকে, X, Y যাদের distribution জানা আছে, তবে তাদের কোনো function-এর distribution-ও বার করে দেওয়া যায়, যেমন $Z = g(X, Y)$-এর distribution. একই কাজ করা যায় যতগুলো খুশী random variable নিয়েও। যেমন ধরো যদি n-খানা random variable নিয়ে শুরু করতাম, $X_1, ..., X_n$, এবং তাদের কোনো একটা function নিতাম, $Y = g(X_1, ..., X_n)$, তাহলে Y-এর distribution বার করা যেত। সেটা করে কী লাভ সেটা বুঝে নিই। ধরো তোমার কাছে একটা random sample আছে $X_1, ..., X_n$. তার ভিত্তিতে তুমি কোনো একটা statistic নিলে $Y = g(X_1, ..., X_n)$. তাহলে আমরা জানতে চাইব যে Y-এ আচরণ কীরকম হবে। তার জন্য Y-এর distribution জানার প্রয়োজন হবে। একটা উদাহরণ নেওয়া যাক।

Example 13: ধরো একটা কয়েন আছে যার $P(\text{head}) = p$. সেটাকে 100 বার টস করে data পাওয়া গেল $X_1, ..., X_{100}$, যেখানে $X_i = 1$ মানে i-th টসে head পড়েছে, নইলে $X_i = 0$. আমরা জানি যে, $\widehat{p} = \sum X_i/100$ হল p-এর একটা estimator. যত বেশী বার টস করা হবে ততই এটা p-এর আরও আরও কাছে এগিয়ে যাবে। প্রশ্ন হল 100 বার টস করার পর $\widehat{p}$-টা p-এর কত কাছে থাকবে? তার জন্য আমাদের এই $\widehat{p}$-এর distribution বার করতে হবে। কী করে বার করতে হবে, সেটা আমরা একটু পরে শিখব। ■

এখানে $\widehat{p}$ ছিল আমাদের random sample, মানে $X_1, ..., X_{100}$-এর একটা function. এরকম function-কে বলা হয় একটা **statistic,** যার সংজ্ঞাটা চেয়েছে নীচের অংকটায়।

Example 14: Define a statistic.[1] **(2009,2004)**

SOLUTION:

> If we have a random sample $X_1, ..., X_n$ then any function $T_n(X_1, ..., X_n)$ is called a <u>statistic</u>. In general, it may depend on the sample size as well as the values of X_i's, but not on any unknown quantity. For example, let $X_1, ..., X_n$ be IID from $N(\mu, 1)$ where μ is unknown. Then $\frac{1}{n}\sum_1^n X_i$ is a statistic, but $\frac{1}{n}\sum_1^n (X_i - \mu)^2$ is not since it involves the unknown μ.

■

Exercise 4: Explain the term: sample characteristics. **(2008,2014)**
HINT: Sample characteristics হল statistic-এরই আরেক নাম। ■

একটা statistic হল random sample-এর function. তাই সেটা নিজেও একটা random variable. অতএব তারও একটা distribution আছে। এই distribution-টাকে বলে statistic-টার **sampling distribution.** সেই নিয়েই এবারের অংকটা।

Example 15: Define sampling distribution of a statistic.[1] **(2009,2005,2004)**

SOLUTION:

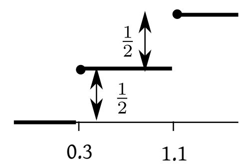


Fig 12

> Let $X_1, ..., X_n$ be a random sample, and $T_n(X_1, ..., X_n)$ be a statistic. Then $T_n(X_1, ..., X_n)$ is also a random variable, and hence has a probability distribution. This distribution is called its sampling distribution.

■

অনেক সময়ে দুটো সম্পূর্ণ আলাদা জিনিসের নামের উচ্চারণ বেশ কাছাকাছি হয়, যেমন '"নাক" আর "টাক"। কিন্তু আসলে এরা এতই আলাদা, যে এদের পার্থক্য লিখতে দিলে ব্যাপারটা হাস্যকর হয়ে পড়ে। তখন এইরকম বোকাবোকা জিনিস লিখতে হয়--টাকে চুল থাকে না, কিন্তু নাকে চুল থাকে। নাক দিয়ে নিঃশ্বাস নেওয়া যায়, কিন্তু টাক দিয়ে যায় না। এবার যে প্রশ্নটা আসছে সেটাও ঠিক এইরকম। এখানে দুটো জিনিস হল sampling distribution আর distribution of a sample. প্রথমটার মানে তো একটু আগেই বললাম। দ্বিতীয়টার মানে হল যাকে আমরা দ্বিতীয় অধ্যায়ে empirical distribution বলেছিলাম। এটা নির্ভর করে sample-টায় কোনো value কবার এসেছে। এদের পার্থক্য লিখতে বলেছে অংকটায়।

Example 16: Distinguish between sampling distribution and distribution of a sample. **(2006.4a)**

SOLUTION:

> Let $X_1, ..., X_n$ be a random sample. Let $T_n(X_1, ..., X_n)$ be a statistic.

Sampling distribution	Distribution of a sample
the probability distribution of $T_n(X_1, ..., X_n)$.	the distribution with CDF $\widehat{F}_n(x) = \frac{1}{n}\lvert\{i \ : \ X_i \leq x\}\rvert$
May be continuous	Must be discrete
Does not depend of a particular sample.	Depends on a particular sample.
Depends on the statistic	Does not depend on any statistic.

মাথা গুলিয়ে যাচ্ছে? একটা উদাহরণ নাও--ধরো, $N(0,1)$ থেকে একটা random sample নিলাম, X_1, X_2. এদের ভিত্তিতে বার করলাম $T_n(X_1, X_2) = X_1 + X_2$. এর distribution দেখানো যায়[1] $N(0, \sqrt{2})$. এইটা হল T_n-এর sampling distribution. যদি $X_1 = 0.3$ আর $X_2 = 1.1$ এসে থাকে, তবে distribution of a sample হবে সেই distribution-টা যার CDF দেখানো হয়েছে Fig 12-এ। এইটাই ছিল প্রথম পয়েন্ট। লক্ষ করো যে, sampling distribution-টা continuous, কিন্তু distribution of the sample হল discrete. এইটা ছিল দ্বিতীয় পয়েন্ট। এবার তৃতীয় পয়েন্ট--যদি X_1, X_2-র value বদলে যায় তবে distribution of the sample-ও বদলে যাবে, কিন্তু sampling distribution-টা একই থাকবে। যদি T_n-এর ফর্মুলা অন্য কিছু নিতাম তবে sampling distribution বদলে যেত, কিন্তু

[1] এটা আমরা কালকে শিখব।

distribution of the sample একই থাকত, কারণ সেখানে T_n ব্যবহারই করা হয়নি। এইটাই ছিল চতুর্থ পয়েন্ট। ■

আমরা যখন কোনো statistic দিয়ে কোনো parameter-কে estimate করি, তখন খানিকটা ভুলচুক হওয়া স্বাভাবিক, কারণ statistic-টা হল random, তাই একটু এদিক ওদিক টলমল করবেই, সেটা তো parameter-এর মত অচল অটল হয়ে একজায়গায় বসে নেই। কতটা টলমল করবে, সেটা মাপা যায় statistic-টার variance (বা তার square root, মানে standard deviation) দিয়ে। এই standard deviation-টার একটা বিশেষ নাম আছে--একে বলে সেই statistic-এর **standard error.**

Example 17: Define standard error of a sample statistic.[1] **(2013)**

SOLUTION: এখানে "sample statistic" বলেছে। খালি "statistic" বললেও একই জিনিস বোঝায়।

> DEFINITION: **Standard error**
> Let T be any sample statistic. Then its standard deviation is called its <u>standard error</u>.

■

31.2 হাতেকলমে

এইবার আমরা হাতেকলমে কিছু sampling distribution বার করব। মানে আমাদেরকে একটা random sample দেওয়া থাকবে, $X_1, ..., X_n$, যাদের distribution আমাদের বলা আছে। এবার কোনো একটা statistic দেওয়া হবে, এবং আমাদের কাজ হবে তার sampling distribution বার করা। কাজটা সহজ নয় মোটেই। আমরা খুব সহজ কয়েকটা ক্ষেত্রেই খালি অংকটা সুন্দর করে করতে পারব, এবং তার বাইরে আমরা যাব না।

31.2.1 Binomial distribution

Example 18: Find the sampling distribution of the mean of a binomial(n, p) variate.[6] **(2012,2008,2004)**

SOLUTION: অংকটা ভুলভাবে লিখেছে। বলেছে mean of a binomial(n, p) variate"। কিন্তু তার মানে হল $E(X)$, যেখানে $X \sim \text{Binomial}(n, p)$. সেটা হল np, যেটা একটা constant, তার আবার sampling distribution কী? আসলে বলতে চেয়েছিল "sample mean of a binomial(n, p) random sample"।
অংকটা বিশেষ রকমের সহজ, তার দুটো কারণ--

- এক, binomial distribution-এর পিছনে একটা সহজ কয়েন টস করার গল্প আছে, সেটা ব্যবহার করা যেতে পারে।
- দুই, binomial distribution-এর characteristic function-টা জটিল কিছু নয়।

আমরা অংকটা তাই দুইভাবে করে দেখাব। প্রথমে খালি কয়েন টসের গল্প দিয়ে।

> Let $X_1, ..., X_m$ be a random sample from $Binomial(n, p)$ distribution.

এখানে একটা ব্যাপারে সাবধান। Binomial distribution-এর প্রথম parameter-টা সাধারণতঃ n লেখাই আমাদের অভ্যাস, আবার sample size-ও সাধারণতঃ আমরা n-ই নিয়ে থাকি। ভুল করে যেন দুটোকেই একই সঙ্গে n নাম দিয়ে বোসো না। আমরা এখানে sample size-টাকে m নাম দিয়েছি।

> We know that a $\text{Binomial}(n, p)$ random variable can be considered as the total number of heads obtained in n independent tosses of a coin with $P(\text{head}) = p$.
>
> So behind each X_i, we have n independent tosses of the same coin.
>
> Thus, there are mn tosses of the coin in total. Since the X_i's are independent, so all these tosses are independent.
>
> Also $X_1 + \cdots + X_m$ is the total number of heads out of all these mn independent tosses of this coin with $P(\text{head}) = p$. So $X_1 + \cdots + X_m$ has $\text{Binomial}(mn, p)$ distribution.

যেহেতু $\sum X_i$-এর distribution পেয়ে গেছি, সেখান থেকেই $\frac{1}{n} \sum X_i$-এর distribution-ও পেয়ে যাব অনায়াসে--

> So the PMF of $\frac{1}{m} \sum X_i$ is
>
> $$P\left(\frac{X_1 + \cdots + X_m}{m} = x\right) = P(X_1 + \cdots + X_m = mx) = \begin{cases} \binom{mn}{mx} p^{mx} (1-p)^{mn-mx} & \text{if } mx \in \{0, 1, ..., mn\} \\ 0 & \text{otherwise} \end{cases}$$

এবার একই অংক করব characteristic function ব্যবহার করে।

বিকল্প পদ্ধতি

> Let $X_1, ..., X_m$ be a random sample from $\text{Binomial}(n, p)$ distribution.
>
> Then, from standard result, each X_k has the same characteristic function
>
> $$Q(t) = (pe^{it} + 1 - p)^n, \quad t \in \mathbb{R}.$$
>
> Since $X_1, ..., X_m$ are independent, hence the characteristic function of $X_1 + \cdots + X_m$ is
>
> $$\underbrace{Q(t) \times \cdots \times Q(t)}_{m \text{ times}} = (pe^{it} + 1 - p)^{mn},$$
>
> which we recognise as the characteristic function of the $\text{Binomial}(mn, p)$ distribution.
>
> Hence, by the characterising property of characteristic function, the distribution of $X_1 + \cdots + X_m$ is $\text{Binomial}(mn, p)$.

এবার $\frac{1}{m} \sum X_i$-এর distribution পেয়ে যেতে আর দেরী নেই--

So the PMF of $\frac{1}{m}\sum X_i$ is

$$\begin{aligned} P\left(\frac{X_1+\cdots+X_m}{m}=x\right) &= P\left(X_1+\cdots+X_m=mx\right) \\ &= \begin{cases} \binom{mn}{mx}p^{mx}(1-p)^{mn-mx} & \text{if } mx\in\{0,1,...,mn\} \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

■

31.2.2 Poisson distribution

এবার একইরকম আরেকটা অংক করব, যেখানে random sample-টা আসছে Poisson distribution থেকে।

Example 19: Find the sampling distribution of the sample mean for the Poisson distribution.

(2014.5c,2011.4d)

SOLUTION:

Let $X_1,...,X_n$ be a random sample from $\text{Poisson}(\lambda)$ distribution.
Want to find the sampling distribution of

$$\overline{X}=\frac{1}{n}\sum_1^n X_i.$$

The characteristic function of $\text{Poisson}(\lambda)$ is

$$Q(t)=\exp[\lambda(e^{it}-1)],\quad t\in\mathbb{R}.$$

Since X_i's are IID $\text{Poisson}(\lambda)$, hence the characteristic function of $\sum_1^n X_i$ is

$$\left(\exp[\lambda(e^{it}-1)]\right)^n=\exp[n\lambda(e^{it}-1)],$$

which is the characteristic function of $\text{Poisson}(n\lambda)$.
So by the characterising property of characteristic functions, $\sum_1^n X_i\sim\text{Poisson}(n\lambda)$.
Hence the sampling distribution of $\overline{X}$ is given by the PMF

$$\begin{aligned} P(\overline{X}=x) &= P\left(\sum_1^n X_i=nx\right) \\ &= \begin{cases} e^{n\lambda}\frac{(n\lambda)^{nx}}{(nx)!} & \text{if } nx\in\{0,1,2,...,\} \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

■

Exercise 5: Find the sampling distribution of the mean of a Poisson μ variate.[5] **(2005.7b)**

HINT:
আগের অংকটাই। তবে ভাষাটা ভুল। “Mean of a Poisson μ variate” বলতে বোঝায় μ, যেটা একটা parameter. তার sampling distribution বলে কোনো কথা হয় না। বলতে চেয়েছিল “sample mean of random sample from Poisson μ distribution”. ■

31.2.3 Finite population

এতক্ষণ আমরা যা যা অংক করেছি, তার সঙ্গে এবারের অংকটার একটা গুরুত্বপূর্ণ পার্থক্য আছে। এতক্ষণ আমরা random sample নিচ্ছিলাম কোনো একটা distribution থেকে, যাকে অনেক সময় infinite population-ও বলা হয়। যদি $X_1, ..., X_n$ এরকম একটা random sample হয়, তবে X_i-রা সবাই IID হবে। কিন্তু এবার আমরা কাজ করব একটা finite population নিয়ে। এখানে ব্যাপারটা অন্যরকম। একটা উদাহরণ দিয়ে বুঝে নিই।

Example 20: ধরো একটা ক্লাসে 10 জন ছাত্রছাত্রী আছে, যাদের রোল নাম্বার 1 থেকে 10. এই হল আমাদের population. এদের থেকে random-ভাবে 3 জনের একটা দল বানাতে হবে। সেটা হবে আমাদের random sample. এর জন্য প্রথমে 10 জনের মধ্য থেকে যে কোনো একজনকে random-ভাবে নেব। ধরো রোল নাম্বার 4-কে নিলাম। এবার অবশিষ্ট $10-1$ জনের থেকে আরেক জনকে নেব random-ভাবে, ধরো এবার পেলাম রোল নাম্বার 7. সবশেষে বাকি $10-2$ জনের থেকে তৃতীয়জনকে নিলাম (ধরো রোল নাম্বার 1)। এইভাবে নেওয়াকে বলে **simple random sample without replacement (SRSWOR)**. এইভাবে পাওয়া তিনজনকে যদি X_1, X_2 আর X_3 বলি (মানে আমাদের উদাহরণে X_1 হল রোল নাম্বার 4, X_2 হল রোল নাম্বার 7, ইত্যাদি) তবে বুঝতেই পারছ যে X_i-রা মোটেই পরস্পর সঙ্গে independent নয় (যেমন, X_1 কোনোভাবেই X_2-র সমান হতে পারে না)। ■

Example 21: Derive the standard error of sample proportion in case of simple random sampling without replacement (SRSWOR).[4] **(2013)**

SOLUTION: এখানে কী করতে হবে সেটা ছাত্রছাত্রীদের উদাহরণ দিয়ে বুঝে নিই।

> Let the population be $\{1, ..., N\}$.

মানে ক্লাসে মোট N জন আছে, যাদের রোল নাম্বার 1 থেকে N পর্যন্ত।

> Let a binary variable y take value $y_r \in \{0, 1\}$ for the r-th population unit, $r = 1, ..., N$.

ধরো, আমরা দেখছি কারা ছেলে আর কারা মেয়ে। ছেলেদের 0 বলব আর মেয়েদের 1. (অন্য কোনো সংখ্যা নিলেও চলত, তবে $0, 1$ নিলে কয়েকটা ধাপ পরে সহজে লেখা যাবে)। সুতরাং "রোল নাম্বার r একজন ছেলে" না বলে বলতে পারি "$y_r = 0$", এইরকম।

> The population proportion is $p = \frac{1}{N}\sum_{r=1}^N y_r$.

অর্থাৎ ক্লাসে মেয়েদের অনুপাত কত। ধরো $N = 5$ জনের মধ্যে 2 জন মেয়ে, রোল নাম্বার 2 আর 4. তার মানে $y_1 = 0,\ y_2 = 1,\ y_3 = 0,\ y_4 = 1,\ y_5 = 0$. সুতরাং মোট মেয়ের সংখ্যা হল $y_1 + \cdots + y_5 = 2$. তার মানে মেয়েদের অনুপাত হচ্ছে $\frac{\sum y_r}{5} = \frac{2}{5}$.

ছোটো ক্লাসের বেলায় এটা বার করা গেল সহজে। কিন্তু যদি পুরো ভারতের population নিয়ে কাজ করতে হত, তবে এই হিসেবটা করা দুঃসাধ্য। তখন population proportion-টাকে unknown parameter হিসেবে estimate করতে হয়। এবং তার জন্য আগে একটা random sample নিতে হয়।

> We draw a SRSWOR of size n. Let the sampled values be $Y_1, ..., Y_n$. Then the sample proportion is
> $$\widehat{p} = \frac{1}{n}\sum_{i=1}^n Y_i.$$

যেমন ধরো, 5 জন ছাত্রছাত্রীর উদাহরণে আমরা $n = 2$ সাইজের একটা SRSWOR নিলাম। যদি এতে রোল নাম্বার 2 আর 5 আসে, তবে সেই দুজনের মধ্যে প্রথমজন মেয়ে আর দ্বিতীয়জন ছেলে, অর্থাৎ $Y_1 = 1$ আর $Y_2 = 0$. সুতরাং আমাদের sample-এ মোট মেয়ের সংখ্যা হল $Y_1 + Y_2 = 1$, তাই মেয়ের অনুপাত হল $\frac{1}{2}$. অন্য একটা SRSWOR নিলে এই অনুপাতটা বদলে যেতে পারে। কতটা বদলে যেতে পারে, সেই বিষয়ে একটা ধারণা পাওয়া যাবে এর variance (বা তার square root, মানে standard error) বার করলে। সেটাই আমরা এই অংকে করতে চলেছি।

> Hence
> $$\text{var}(\widehat{p}) = \text{var}\left(\frac{1}{n}\sum_{i=1}^n Y_i\right) = \frac{1}{n^2}\left[\sum_{i=1}^n \text{var}(Y_i) + \sum_{i\neq j}^n \text{cov}(Y_i, Y_j)\right].$$

তার মানে আমাদের লাগবে $\text{var}(Y_i)$ আর $\text{cov}(Y_i, Y_j)$. এগুলো একে একে বার করা যাক। প্রথমে $\text{var}(Y_i)$.

> Now for each $i \in \{1, ..., n\}$, the random variable Y_i can take values $y_1, ..., y_N$ each with probability $1/N$. So
> $$E(Y_i) = \sum_{r=1}^N y_r \times \frac{1}{N} = p, \text{ and } E(Y_i^2) = E(Y_i) = p.$$
> So
> $$\text{var}(Y_i) = E(Y_i^2) - \big(E(Y_i)\big)^2 = p - p^2 = p(1-p).$$

এবার $\text{cov}(Y_i, Y_j)$ বার করব--

> Also, for $i \neq j \in \{1, ..., n\}$, the pair (Y_i, Y_j) can take the values (y_r, y_s)

$(r \neq s)$ with equal probability, $1/(N(N-1))$. So

$$\begin{aligned}
E(Y_iY_j) &= \frac{1}{N(N-1)}\sum_{r=1}^{N}\sum_{s\neq r}^{N} y_r y_s \\
&= \frac{1}{N(N-1)}\sum_{r=1}^{N}\left[y_r \sum_{s\neq r}^{N} y_s\right] \\
&= \frac{1}{N(N-1)}\sum_{r=1}^{N}[y_r(Np-y_r)] \\
&= \frac{1}{N(N-1)}\left[Np\sum_{r=1}^{N} y_r - \sum_{r=1}^{N} y_r^2\right] \\
&= \frac{1}{N(N-1)}\left[Np\sum_{r=1}^{N} y_r - \sum_{r=1}^{N} y_r\right] \quad [\because y_r \in \{0,1\}, \quad \therefore y_r^2 = y_r] \\
&= \frac{(Np-1)Np}{N(N-1)} \\
&= \frac{(Np-1)p}{N-1}
\end{aligned}$$

So

$$\begin{aligned}
\operatorname{cov}(Y_i, Y_j) &= E(Y_iY_j) - E(Y_i)E(Y_j) \\
&= \frac{(Np-1)p}{N-1} - p^2 \\
&= -\frac{p(1-p)}{N-1}.
\end{aligned}$$

ব্যস্, $\operatorname{var}(Y_i)$ আর $\operatorname{cov}(Y_i, Y_j)$ বেরিয়ে গেছে। সুতরাং এবার $\operatorname{var}(\widehat{p})$ বার করে ফেলতে আর সমস্যা নেই--

Hence

$$\begin{aligned}
\operatorname{var}(\widehat{p}) &= \frac{1}{n^2}\left[\sum_{i=1}^{n}\operatorname{var}(Y_i) + \sum_{i\neq j}^{n}\operatorname{cov}(Y_i, Y_j)\right] \\
&= \frac{1}{n^2}\left[\sum_{i=1}^{n} p(1-p) - \sum_{i\neq j}^{n}\frac{p(1-p)}{N-1}\right] \\
&= \frac{1}{n^2}\left[np(1-p) - \frac{n(n-1)p(1-p)}{N-1}\right] \\
&= \cdots = \frac{p(1-p)}{n} \times \frac{N-n}{N-1}.
\end{aligned}$$

■

DAY 32
Sampling distributions related to normal samples

আজকে আমরা আরও sampling distribution বার করব। এবার। random sample-টা আসবে কোনো একটা normal distribution থেকে। গতকাল আমরা খালি একটা statistic নিয়ে কাজ করছিলাম, sample mean. এবার কাজ করব দুটো statistic নিয়ে। Sample mean তো থাকবেই, তার সঙ্গে sample variance-কেও হিসেবে নেব।
এই অংকগুলোর জন্য characteristic function ব্যবহার করলে খুবই সুবিধা হবে। চট্ করে মনে করে নিই $N(\mu,\sigma)$ distribution-এর characteristic function কী ছিল--

$$Q(t) = e^{it\mu - \frac{1}{2}t^2\sigma^2}.$$

তার মানে যদি কখনও দ্যাখো যে কোনো random variable-এর characteristic function-এর চেহারা এইরকম--

$$\exp(it \times \text{ কোনো সংখ্যা} - t^2 \times \text{ আরেকটা সংখ্যা}),$$

তবে জানবে যে random variable-টার distribution অবশ্যই normal হতে বাধ্য, এবং ওই "কোনো সংখ্যা"-টাই ওর mean, আর "আরেকটা সংখ্যা"-টার ডবল হল ওর variance.
এই বার তবে প্রথম অংকে হাত দেওয়া যাক।

Example 22: Find the sampling distribution of the sample mean

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

of a sample $x_1, ..., x_n$ of size n drawn from a normal population with mean m and standard deviation σ.[7] **(2009)**
SOLUTION:

Let $X_1, ..., X_n$ be a random sample from $N(\mu,\sigma)$ $(n \geq 1)$.
Then each X_k has characteristic function

$$Q(t) = e^{it\mu - \frac{t^2\sigma^2}{2}}.$$

Also the X_k's are all independent. So the characteristic function of $\bar{X}$ is

$$\begin{aligned} E(e^{it\bar{X}}) &= E\left(\exp\left[it\tfrac{1}{n}(X_1 + \cdots + X_n)\right]\right) \\ &= E\left(\exp\left[i\tfrac{t}{n}X_1\right] \times \cdots \times \exp\left[i\tfrac{t}{n}X_n\right]\right) \\ &= E\left(\exp\left[i\tfrac{t}{n}X_1\right]\right) \times \cdots \times E\left(\exp\left[i\tfrac{t}{n}X_n\right]\right) \quad [\because \text{independent}] \\ &= \underbrace{Q\left(\tfrac{t}{n}\right) \times \cdots \times Q\left(\tfrac{t}{n}\right)}_{n \text{ times}} \\ &= \left(e^{-i\frac{t}{n}\mu - \frac{t^2\sigma^2}{2n^2}}\right)^n \\ &= e^{-it\mu - \frac{t^2\sigma^2}{2n}}, \end{aligned}$$

which we recognise as the characteristic function of the $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ dis-

tribution.
So, by the characterising property of characteristic function, the sampling distribution of $\bar{X}$ is $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

■

ধাপগুলো দেখতে একটু লম্বা বটে, কিন্তু মূল ধারণাটা কিন্তু তেমন শক্ত কিছু নয়। পুরো ব্যাপারটাকে একটা theorem-এর আকারে মনে রাখলে সুবিধা হবে--

THEOREM

Let $X_1, ..., X_n$ be IID $N(\mu, \sigma)$. Let the sample mean be

$$\overline{X} = \frac{1}{n}\sum_1^n X_i.$$

Then

$$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

এবার এর একটা প্রয়োগ দেখি।

Example 23: The variable X is normally distributed with mean 68 cm and s.d. 2.5 cm. What should be the size of the sample whose mean shall not differ from the population mean by more than 1 cm with probability 0.95?[4] **(2011.4c)**

SOLUTION: কী চাইছে বুঝে নিই। একটা random variable আছে X, তার mean বলেই দিয়েছে 68 cm. যেহেতু random variable, তাই এর পিছনে একটা random experiment আছে, যেটাকে বারবার করে করলে X-এর বিভিন্ন valueপাওয়া যাবে। ধরো n-বার করলাম। এর ফলে X-এর যে n-খানা value পেলাম তাদের average করলে পাবো sample mean $\bar{X}$. এটা 68 cm-এর ধারেকাছেই কোথাও থাকবে সেটা প্রত্যাশিত। কিন্তু এমন কী জোর দিয়ে বলা যায় যে, $\bar{X}$ নিশ্চয়ই 68 ± 1 cm-এর মধ্যে থাকবেই? না, সেটা বলা যায় না বটে, কিন্তু n যতই পারবে ততই $\bar{X}$-এর পক্ষে 68 ± 1 cm-এর মধ্যে থাকার probability বাড়বে। এই অংকটায় জানতে চেয়েছে n ঠিক কতটা বড় হলে এই probability-টা 0.95 হবে।

Let n be the required sample size.
Let $X_1, ..., X_n$ be iid $N(68, 2.5)$.
Then from standard result

$$\bar{X} = \frac{1}{n}\sum_1^n X_i \sim N\left(68, \frac{2.5}{\sqrt{n}}\right).$$

So

$$P(|\bar{X} - 68| \leq 1) = P\left(\frac{|\bar{X}-68|}{2.5/\sqrt{n}} \leq \frac{1}{2.5/\sqrt{n}}\right) = 2\left(\Phi\left(\frac{1}{2.5/\sqrt{n}}\right) - \frac{1}{2}\right),$$

> where $\Phi(x)$ is the $N(0,1)$ CDF.

এই জায়গাটা বুঝলে তো? যেহেতু $\bar{X} \sim N\left(68, \frac{2.5}{\sqrt{n}}\right)$, তাই $\frac{\bar{X}-68}{2.5/\sqrt{n}} \sim N(0,1)$.

> We want this to be ≥ 0.95,

এখানে আমরা $= 0.95$ না লিখে ≥ 0.95 লিখলাম কেন? কারণ n যেহেতু ভগ্নাংশ হতে পারে না, তাই একেবারে probability-টাকে একেবারে $= 0.95$ বানানো নাও যেতে পারে। তাই আমরা চাই যেন অন্ততঃ 0.95 গ্যারান্টি দেওয়া যায়।

> ie,
>
> $$2\left(\Phi\left(\frac{1}{2.5/\sqrt{n}}\right) - \frac{1}{2}\right) \geq 0.95,$$
>
> or
>
> $$\Phi\left(\frac{1}{2.5/\sqrt{n}}\right) \geq 0.975,$$
>
> or
>
> $$\frac{\sqrt{n}}{2.5} \geq \Phi^{-1}(0.975),$$
>
> since $\Phi(x)$ is a strictly increasing function.
>
> Thus
>
> $$n \geq \left(2.5\Phi^{-1}(0.975)\right)^2 = 24.009.$$

এটার জন্য তোমার জানা থাকতে হবে যে, $\Phi^{-1}(0.975) = 1.96$.

> So the sample size should be $n \geq 25$.

কারণ n তো আর ভগ্নাংশ হতে পারে না। ■

এবার আমরা sample variance-এর sampling distribution বার করার দিকে এগোব। তার জন্য আমাদের একটা নতুন probability distribution-এর কথা জানতে হবে--

DEFINITION: χ^2 distribution

If $X_1, ..., X_n$ are iid $N(0,1)$, then the distribution of the random variable

$$X_1^2 + \cdots + X_n^2$$

is called $\chi^2_{(n)}$ (chi-square distribution with n degrees of freedom).

এর characteristic function-এর সঙ্গে পরিচয় থাকা ভালো, কারণ এক্ষুণি কাজে লাগবে--

THEOREM

The characteristic function of $\chi^2_{(k)}$ distribution is $(1-2it)^{-k/2}$.

এবার আমরা যেটা প্রমাণ করতে চলেছি সেটাকে একটা theorem-এর আকারে লিখে রাখলে সুবিধা হবে--

THEOREM

Let $X_1, ..., X_n$ be IID $N(\mu, \sigma)$, where $n \geq 2$. Let the sample variance be

$$S^2 = \frac{1}{n} \sum_1^n (X_i - \overline{X})^2.$$

Then

$$\frac{S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

এই theorem-টা প্রমাণ করতে বলেছে এর পরের অংকটায়। প্রমাণটার জন্য আবার আরেকটা theorem লাগবে, সেটা লিখে রাখি--

THEOREM

Let $X_1, ..., X_n$ be IID $N(\mu, \sigma)$. Let $Y_1, ..., Y_p$ be defined as

$$\begin{aligned} Y_1 &= a_{11}X_1 + \cdots + a_{1n}X_n \\ &\vdots \\ Y_p &= a_{p1}X_1 + \cdots + a_{pn}X_n. \end{aligned}$$

Also let Z be defined as

$$Z = b_1 X_1 + \cdots + b_n X_n.$$

Then $(Y_1, ..., Y_p)$ are independent of Z if and only if

$$\forall i \in \{1, ..., p\} \quad \text{cov}(Z, Y_i) = 0.$$

এই theorem-এর প্রমাণটা এখানে করব না। কিন্তু theorem-এর মূল বক্তব্যটা বুঝে রাখা যাক। ধরো, দুটো random variable আছে, X আর Y, যাদের covariance-এর অস্তিত্ব আছে। আমরা জানি যে, X, Y যদি independent হয়, তবে $\text{cov}(X, Y) = 0$ হতে বাধ্য, কিন্তু উল্টো দিকটা জোর দিয়ে বলা যায় না, মানে $\text{cov}(X, Y) = 0$ হলেও X, Y কিন্তু independent নাও হতে পারে। এই theorem-টা বলেছে যে, normal sample-এর বেলায় উল্টো দিকটাও সত্য।

Example 24: For a normal(m, s) population show that the statistic

$$\frac{nS^2}{s^2}$$

is χ^2-distributed with $(n-1)$ degrees of freedom, where n, S^2 and s^2 are sample size, sample variance and population variance, respectively.[5] **(2004.7b)**

SOLUTION: এখানে statistic শব্দটা সঠিকভাবে ব্যবহার করা হয় নি। Statistic মানে এমন কোনো জিনিস যেটা কেবলমাত্র random sample-টার function, মানে যার মধ্যে কোনো অজানা parameter নেই। এখানে আমরা জানিনা s-টা জানা নাকি অজানা। যেহেতু $\frac{nS^2}{s^2}$-এর মধ্যে s^2 ঢুকে আছে, তাই ওকে statistic বলা উচিত নয়।
অংকটা একটু কৌশলে করব, কয়েকটা ধাপে ভেঙে--

> Step 1: Shall show that
>
> $$\sum(X_i-m)^2=\sum(X_i-\bar{X})^2+(\bar{X}-m)^2.$$
>
> Here
>
> $$\begin{aligned} LHS &= \sum(X_i-m)^2 \\ &= \sum(\underbrace{X_i-\bar{X}}+\underbrace{\bar{X}-m})^2 \end{aligned}$$
>
> এবার $(a+b)^2$-এর ফর্মুলা লাগাও। আর মনে রেখো যে $\sum(X_i-\bar{X})=0$.
>
> $$\begin{aligned} &= \ldots \\ &= \sum(X_i-\bar{X})^2+(\bar{X}-m)^2=RHS. \end{aligned}$$

এইবার দ্বিতীয় ধাপে আমরা সেই theorem-টা লাগাব।

> Step 2: Shall show that $X_i-\bar{X}$ is independent of $\bar{X}$.

Theorem-টার কল্যাণে এর জন্য খালি এদের covariance-টা শূন্য দেখাতে পারলেই হবে--

> $$\begin{aligned} \mathrm{cov}(X_i-\bar{X},\bar{X}) &= \mathrm{cov}(X_i,\bar{X})-\mathrm{cov}(\bar{X},\bar{X}) \\ &= \mathrm{cov}(X_i,\bar{X})-\mathrm{var}(\bar{X}) \\ &= \tfrac{1}{n}\mathrm{cov}(X_i,X_i)-\mathrm{var}(\bar{X}) \\ &= \tfrac{1}{n}\mathrm{var}(X_i)-\mathrm{var}(\bar{X}) \\ &= 0. \end{aligned}$$
>
> So, by a standard result of normal samples, $X_i-\bar{X}$ is independent of $\bar{X}$.

এখানে standard result মানে ওই theorem-টা।
এবার characteristic function-এর খেল শুরু হবে--

> Step 3: We know that if U has a $N(\mu,\sigma)$ distribution, then $\frac{U-\mu}{\sigma}$ has a $N(0,1)$ distribution, and hence $\left(\frac{U-\mu}{\sigma}\right)^2$ has a $\chi^2_{(1)}$ distribution.

এটা χ^2-distribution-এর সংজ্ঞা থেকেই আসছে।

From step 1,

$$\sum \left(\frac{X_i - m}{s}\right)^2 = \frac{1}{s^2}\sum (X_i - \bar{X})^2 + \left(\frac{\bar{X} - m}{s}\right)^2.$$

Let the three terms have characteristic function's $Q_1(t), Q_2(t)$ and $Q_3(t)$, respectively.

By step 2, the two terms in the RHS are independent, and hence

$$Q_1(t) = Q_2(t)Q_3(t).$$

Now we are given that $X_i \sim N(m, s)$ and we know that $\bar{X} \sim N(m, s/\sqrt{n})$.

So the LHS term has $\chi^2_{(n)}$ distribution, and the last term in the RHS has a $\chi^2_{(1)}$ distribution.

So $Q_1(t) = (1 - 2it)^{-n/2}$ and $Q_3(t) = (1 - 2it)^{-1/2}$.

Thus

$$Q_2(t) = Q_1(t)/Q_3(t) = (1 - 2it)^{-(n-1)/2},$$

which we recognise as the characteristic function of $\chi^2_{(n-1)}$ distribution.

So, by the characterising property of characteristic function,

$$\frac{nS^2}{s^2} \sim \chi^2_{(n-1)},$$

as required.

■

Exercise 6: For a normal(m, σ) population, show that the statistic $\frac{ns^2}{\sigma^2}$ has χ^2-distribution with $(n-1)$ degrees of freedom, where n, s^2 and σ^2 are sample size, sample variance and population variance, respectively.[7] **(2011.4b)**

HINT:

আগের অংকটাই। ■

এবার একটা জিনিস লক্ষ কর। আমরা অংকটা তিনটে ধাপে ভেঙে করেছিলাম। তার মধ্যে দ্বিতীয় ধাপে আমরা যেটা দেখিয়েছি সেটা থেকে একটা দরকারী তথ্য পাওয়া যায়--normal sample-এর ক্ষেত্রে sample mean আর sample variance পরস্পরের সঙ্গে independent. এই কথাটা আমাদের এক্ষুণি কাজে লাগবে। তাই একটা theorem-এর আকারে লিখে রাখি--

THEOREM

Let $X_1, ..., X_n$ be IID $N(\mu, \sigma)$, where $n \geq 2$. Let the sample mean and sample variance be, respectively,

$$\overline{X} = \frac{1}{n}\sum_1^n X_i \text{ and } S^2 = \frac{1}{n}\sum_1^n (X_i - \overline{X})^2.$$

Then $\overline{X}$and $S^2$are independent.

এবার আমাদের আলোচনাটা যেদিকে যাবে, তার জন্য একটা নতুন distribution-এর সংজ্ঞা জানতে হবে--

DEFINITION: t distribution

If $X \sim N(0,1)$ and $Y \sim \chi^2_{(k)}$ are independent, then the distribution of the random variable

$$\frac{X}{Y/\sqrt{k}}$$

is called $t_{(n)}$ (t distribution with n degrees of freedom).

Example 25: Let $x_1, ..., x_n$ be a random sample of size n $(n > 1)$ from a normal(m, σ) population. Find the sampling distribution of the statistic

$$t = \sqrt{n}(\overline{x} - m)/s,$$

where $\overline{x}$ is the sample mean, and s^2 is the unbiased estimate of population variance σ^2.[7] **(2007.4b)**

SOLUTION: এই অংকে বোঝাতে চেয়েছে যে,

$$s^2 = \frac{1}{n-1}\sum(x_i - \overline{x})^2.$$

কিন্তু প্রশ্নের ভাষাটা ঠিক হয়নি, কারণ "the unbiased estimate of population variance" বলে কিছু হয় না। Unbiased estimator মোটেই unique কিছু নয় যে, σ^2-এর unbiased estimate মানেই এই s^2-টা বুঝতে হবে। উত্তরের শুরুতেই আমাদের notation ঠিক করে নিই--

> Let $X_1, ..., X_n$ be a random sample from $N(m, \sigma^2)$ $(n \geq 2)$.
> Let
>
> $$\overline{X} = \frac{1}{n}\sum X_i$$
> $$S^2 = \sum(X_i - \overline{X}_n)^2$$

Random variable-দের বড়হাতের ইংরাজি অক্ষর দিয়ে বোঝানোই প্রচলিত নিয়ম। তাই প্রশ্নে ছোটোহাত দেওয়া থাকলেও আমরা বড়হাত করে নিয়েছি।

এবার আগের অংক দুটোর সমাধান এখানে পুরো লিখতে হবে, মানে $\overline{X}$-এর distribution বার করে দেখাতে হবে, S^2-এর distribution বার করে দেখাতে হবে, এবং $\overline{X}$ আর S^2 যে পরস্পরের সঙ্গে independent, সেটাও দেখাতে হবে। (লিখতে লিখতে আঙুল খুলে যাবার জোগাড় হবে!) সবশেষে উপসংহারটা হবে এইরকম--

Let

$$X = \frac{\overline{X} - m}{\sigma/\sqrt{n}},$$
$$Y = \frac{1}{\sigma^2}\sum_1^n (X_i - \overline{X})^2.$$

Then we have shown that

1. $X \sim N(0,1)$,
2. $Y \sim \chi^2_{(n-1)}$,
3. X, Y are independent.

So, by definition of t-distribution,

$$t = \frac{\sqrt{n}(\overline{X} - m)}{S} = \frac{X}{\sqrt{Y/(n-1)}} \sim t_{(n-1)},$$

which is the required distribution.

■

Exercise 7: Let $X_1, X_2, ..., X_n$ be a random sample of size $n(\geq 1)$ from a $N(m, \sigma)$ population. Find the sampling distribution of the statistic

$$t = \frac{(\overline{x} - m)\sqrt{n}}{s}$$

where $\overline{x}$ is the sample mean and

$$(n-1)s^2 = \sum_{i=1}^n (x_i - \overline{x})^2.$$

[6] **(2005.7c)**
HINT:
আগের অংকটাই। ■

Exercise 8: Let $X_1, X_2, ..., X_n$ be a random sample of size $n(> 1)$ from a normal(m, σ) population. Find the sampling distribution of the statistic

$$t = \frac{\sqrt{n}(\bar{X} - \mu)}{S},$$

where $\bar{X}$ is the sample mean and S^2 is the unbiased estimate of population variance σ^2. **(2010.4a)**

HINT:
আবারও আগের অংকটাই। ■

এবারের অংকটা একেবারেই অন্যরকম। এটা আসলে MLE-র অংক, খালি distribution-টা একটু ঘুরিয়ে দিয়েছে বলে এখানে আলোচনা করছি। এই অংকটার জন্য $\chi^2_{(k)}$ distribution-এর PDF-এর চেহারাটা কাজে লাগবে--

> **THEOREM**
>
> The $\chi^2_{(k)}$ distribution has PDF
>
> $$f(x) = \begin{cases} \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

গোড়ার ওই বিচ্ছিরি Γ ইত্যাদি দেখে ভয় পেও না, ওগুলো এই অংকে লাগবে না। খালি PDF-টার আদলটুকু মনে রাখলেই চলবে।

Example 26: Let $s_1^2, .., s_k^2$ denote a sample of size k from the population of the unbiased estimate of the population variance for a normal(m, σ) population. Write down the likelihood function of this sample, and show that maximum likelihood estimate of σ^2 is given by

$$\widehat{\sigma}^2 = (s_1^2 + \cdots + s_k^2)/k.$$

[7] **(2007.4c)**
SOLUTION: প্রশ্নটার ভাষা ভুল এবং বোঝাও কঠিন। ভুলটা আগের অংকের ভাষাতেও ছিল-- "the unbiased estimate of population variance" বলে কিছু হয়না। আসলে বলতে চেয়েছিল যে, $s_1^2, ..., s_k^2$ IID $\chi^2_{(n-1)}$. সেটাকেই ঘুরিয়ে কঠিন করে বলবার চেষ্টা করেছে।

> We consider the unbiased estimates
>
> $$S^2 = \frac{1}{n-1}\sum_1^n (X_i - \overline{X})^2$$
>
> based on samples of size n.
> We assume that $S_1^2, ..., S_k^2$ are IID with same distribution as S^2.
> So, using standard result, we know that
>
> $$Y_i = \frac{n-1}{\sigma^2} S_i^2 \sim \chi^2_{(n-1)}.$$
>
> So the pdf of Y_i is
>
> $$f(y) = \text{const} \times e^{-\frac{y}{2}} y^{\frac{n-1}{2}-1}, \quad y > 0,$$

where "const" is a known constant.

লক্ষ কর আমরা কীভাবে χ^2 PDF-এর বিচ্ছিরি constant-টাকে স্রেফ const লিখে ছেড়ে দিয়েছি। Constant-টা ঠিক কী, সেটা আমাদের অংকে মোটেই লাগবে না।
এইবার আমরা S_i^2-দের PDF বার করব Jacobian formula লাগিয়ে। এখানে parameterহবে σ^2. S_i^2 আর σ^2 দুটোতেই square রয়েছে, যেগুলো লেখার পথে অসুবিধা সৃষ্টি করবে, তাই একটু নতুন নাম দিয়ে নিই--

Let $Z_i = s_i^2$ and $\theta = \sigma^2$.
Then by Jacobian formula the PDF of $Z_i = \frac{\theta}{n-1} Y_i$ is

$$g(z) = \frac{n-1}{\theta} f\left(\frac{n-1}{\theta} z\right)$$

এখানে Jacobian-টা হল $\frac{n-1}{\theta}$.

$$= \text{const} \times \frac{1}{\theta} \exp\left(-\tfrac{n-1}{2\theta} z\right) \left(\frac{n-1}{\theta} z\right)^{\frac{n-3}{2}}$$

$$= \text{const} \times \theta^{-\frac{n-1}{2}} \exp\left(-\tfrac{n-1}{2\theta} z\right) z^{\frac{n-3}{2}}$$

So the likelihood function is

$$L(\theta) = \prod g(z_i)$$

$$= \text{const} \times \theta^{-\frac{k(n-1)}{2}} \exp\left(-\tfrac{n-1}{2\theta} \sum z_i\right) \left(\prod z_i\right)^{\frac{n-3}{2}}$$

So the log-likelihood function is

$$\ell(\theta) = \text{const} - \tfrac{k(n-1)}{2} \log\theta - \tfrac{n-1}{2\theta} \sum z_i + \tfrac{n-3}{2} \sum \log z_i.$$

So

$$\ell'(\theta) = -\tfrac{k(n-1)}{2\theta} + \tfrac{n-1}{2\theta^2} \sum z_i.$$

Equating to 0 we get

$$\tfrac{k(n-1)}{2\theta} = \tfrac{n-1}{2\theta^2} \sum z_i,$$

or

$$\widehat{\theta} = \tfrac{1}{k} \sum z_i.$$

The second derivative is

$$\ell''(\theta) = \tfrac{k(n-1)}{2\theta^2} - \tfrac{n-1}{\theta^3} \sum z_i.$$

> So
>
> $$\begin{aligned} \ell''(\widehat{\theta}) &= \tfrac{k(n-1)}{2\widehat{\theta}^2} - \tfrac{n-1}{\widehat{\theta}^3}\sum z_i \\ &= \tfrac{1}{\widehat{\theta}^2}\left(\tfrac{k(n-1)}{2} - \tfrac{n-1}{\widehat{\theta}}\sum z_i\right) \\ &= \tfrac{1}{\widehat{\theta}^2}\left(\tfrac{k(n-1)}{2} - k(n-1)\right) < 0. \end{aligned}$$
>
> So $\ell(\theta)$ is maximised at $\widehat{\theta}$.

এবার উত্তরটাকে প্রশ্নের notation-এ লিখব মানে Z_i না লিখে S_i^2 এবং θ-র বদলে σ^2.

> So the required MLE of σ^2 is
>
> $$\widehat{\sigma}^2 = \frac{1}{k}\sum_1^k S_i^2.$$

■

Answers

1. $\begin{cases} 1 - e^{-\sqrt{y}} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$ **2.** $aX + b \sim N(am + b, |a|\sigma)$.

Chapter IX
Properties of estimators and confidence intervals

DAY 33
Properties of estimators

Estimation করা বলতে কী বোঝায়, তা আমরা দ্বিতীয় অধ্যায়েই শিখেছি। ব্যাপারটাকে ভগবানের সঙ্গে statistician-এর খেলা বলে ভাবা যায়। ভগবানের কাছে যেন একটা random experiment আছে, সেটা থেকে যে random variable বেরোয় তাদের distribution-টা তোমার জানা নেই। কিন্তু distribution-টা কোন family-তে আছে, সেটা তুমি আন্দাজ করে ফেলেছ। সেই family-র বিভিন্ন সদস্যদেরকে আলাদা করার জন্য আছে parameter, ধরো θ. তুমি family-টা আন্দাজ করে ফেলেছ ঠিকই, কিন্তু সেই family-র ঠিক কোন সদস্যকে ভগবান ব্যবহার করেছেন সেটা আন্দাজ করা এখনও বাকি আছে, মানে θ আন্দাজ করা। আমাদের হাতে আছে data, $X_1, ..., X_n$. সেই data-এর ভিত্তিতে θ আন্দাজ করতে পারাই হল estimation. তার একটা কায়দা আমরা ইতিমধ্যেই শিখেছি--maximum likelihood estimation. ধরো তুমি সেই কায়দায় (বা অন্য কোনো কায়দায়) θ-র value-টা আন্দাজ করেছ। এই আন্দাজটাকে লিখব $\widehat{\theta}_n(X_1, ..., X_n)$, বা সংক্ষেপে খালি $\widehat{\theta}_n$. এখানে নীচে n লিখেছি এইটা মনে করিয়ে দেওয়ার জন্য যে, আমাদের sample size হল n. এই $\widehat{\theta}_n$-কে আমরা বলব θ-র একটা estimator. যেমন যদি $X_1, ..., X_n$ হয় IID $N(\theta, 1)$, তবে θ-র MLE আমরা বার করেছিলাম $\overline{X}$. তার মানে সেখানে $\widehat{\theta}_n = \overline{X}$.

এবার স্বভাবতঃই প্রশ্ন উঠবে যে, θ-র estimator হিসেবে $\widehat{\theta}_n$ কতটা ভালো। সেটা বোঝার একটা কায়দা হল ভগবানকে চুপিচুপি জিজ্ঞেস করা--"দাদা[1], আপনি θ-র ঠিক কি value নিয়েছিলেন গো?" ভগবান যদি তোমার কাতর আকুতিতে সাড়া দেন তবে $\widehat{\theta}_n$-এর সঙ্গে সেই প্রকৃত value-র তুলনা করলেই গোল মিটে যাবে। তবে কিনা ভগবানকে একটা কাতর আকুতি করলেই যদি তিনি parameter-এর value-টা বলে দেন, তবে আর এত কষ্ট করা কেন বাপু? গোড়াতেই ভগবানকে জিজ্ঞেস করে নিলেই তো ল্যাঠা চুকে যেত! যে জিনিসটা এত সহজেই জানা যায়, তার জন্য statistics-এর ধুন্ধুমারের দরকার নেই। নইলে ব্যাপারটা "শোলে" সিনেমার মত হয়ে যাবে, যেখানে একজন প্রশ্ন করেছিল "তুমহারা নাম কেয়া হ্যায়, বাসন্তী?"

কিন্তু কথা হচ্ছে এই যে, ভগবান সাধারণতঃ এই ধরণের কাতর আবেদনে সাড়া দেন না। তাই parameter-এর value-ও অজানাই থাকে। আর তাই গোড়ার প্রশ্নটাও থেকেই যায়-- প্রকৃত value-টা না জেনে কী করে বিচার করা যাবে যে, estimator-টা কত ভালো হয়েছে?

এইরকম বিচার করার বিভিন্ন পদ্ধতি আছে, যার মধ্যে দুটো হল unbiasedness আর consistency, যেগুলো এবার আমরা আলোচনা করব।

33.1 Unbiasedness এবং consistency

[1] অবশ্য তোমার বিশ্বাস অনুযায়ী তুমি "দিদি" বা "মাসী" বলেও ডেকে দেখতে পারো। সম্ভবতঃ একই ফল পাবে।

> তাঁর বিচারের নিয়মটি কিন্তু ভারি সহজ। ... যাদের উপরে তিনি খুব খুশি থাকেন, যারা তাঁকে পয়সা-টয়সা দেয়, ... তাদের তলোয়ার দেখলেই তিনি নেড়েচেড়ে টিপেটুপে বলেন, "খাসা তলোয়ার! ... হাজার টাকা দামের তলোয়ার!" আর যাদের উপরে তিনি চটা,... তাদের তলোয়ার ... হাতে পড়লেই তিনি অমনি একটু শুঁকেই নাক সিঁট্‌কিয়ে বলে ওঠেন, "অতি বিচ্ছিরি! ... তলোয়ার তো নয়, যেন কাস্তে গড়েছে!"
>
> --অসিলক্ষণ পণ্ডিত (সুকুমার রায়)

মনে রেখো estimator-টা একটা random variable, কারণ সেটা data-র উপর নির্ভর করে, এবং data হল random. ওদিকে parameter-টা একটা constant. অতএব estimator-টা যে একেবারে গ্যারান্টি দিয়ে parameter-এর সমান হবে, এমনটা আশা করাই অন্যায়। মোটামুটি কাছাকাছি গেলেই আমরা খুশি। এই "মোটামুটি কাছাকাছি যাওয়া"-র ব্যাপারটাকেই আমরা অংকের ভাষায় প্রকাশ করতে চাই। এক কথায় করা মুশকিল, ভেঙে ভেঙে এগোই।
একটা অজানা parameter হল θ. তাহলে $E(\widehat{\theta}_n)$ থেকে বোঝা যাবে যে estimator-টা মোটামুটি কোন value-র আশেপাশে ঘোরাঘুরি করে। যদি $E(\widehat{\theta}_n)$-টা প্রকৃত value-র সমান হয় তবে আমরা খুশি। সেক্ষেত্রে আমরা বলব যে, $\widehat{\theta}_n$ হল θ-র একটা unbiased estimator. এখানে একটা ব্যাপার আছে, আমরা তো আমাদের data-র distribution-টা সম্পূর্ণ জানিনা, খালি জানি distribution-টা কোনো family-তে আছে। এই family-র বিভিন্ন সদস্য distribution-এর বেলায় $E(\widehat{\theta}_n)$ আলাদা আলাদা হতেই পারে। Unbiased হবার জন্য তাই আমরা চাই যেন এই family-র যেকোনো সদস্য distribution-এর জন্যই $E(\widehat{\theta}_n) = \theta$ হয়।

Example 1: Define an unbiased estimate of a parameter of a population distribution.[2] **(2004,2005,2007,2009,2012)**

SOLUTION:

> DEFINITION: **Unbiased**
>
> Let $X_1, ..., X_n$ be a random sample from a distribution belonging to a known family. Let θ be an unknown parameter, and let $\widehat{\theta}_n(X_1, ..., X_n)$ be an estimator of θ. We say that $\widehat{\theta}_n(X_1, ..., X_n)$ is unbiased for θ, if for all distributions in the family
>
> $$E(\widehat{\theta}_n(X_1, ..., X_n)) = \theta.$$
>
> An estimator that is not unbiased is called biased.

■

Example 2: Prove that the maximum likelihood estimate of the parameter α of the population having density function

$$f(x) = \frac{2(\alpha - x)}{\alpha^2}, \quad (0 < x < \alpha),$$

for a sample x_1 of unit size is $2x_1$, and that this estimate is biased.[6] **(2011)**

SOLUTION: অংকটার প্রথম অংশ (মানে MLE প্রমাণ করা অব্ধি) আমরা আগেই করেছি (দ্বিতীয় অধ্যায়ে)। এবার biased প্রমাণ করব।

> We shall show that $E(2X_1) \neq \alpha$ for at least one value of α.
>
> $$\begin{aligned} E(2X_1) &= \int_{-\infty}^{\infty} 2xf(x)dx \\ &= \int_0^{\alpha} \frac{4x(\alpha - x)}{\alpha^2} dx \\ &= 4\int_0^{\alpha} \tfrac{x}{\alpha}\left(1 - \tfrac{x}{\alpha}\right) dx \\ &= 4\alpha \int_0^1 u(1-u)du \quad \left[\text{putting } u = \tfrac{x}{\alpha}\right] \\ &= \tfrac{2\alpha}{3}. \end{aligned}$$
>
> Thus, $E(2X_1) \neq \alpha$ for any $\alpha > 0$.
>
> Hence the MLE is biased for α, as required.

■

আমরা শিখছিলাম কীভাবে কোনো estimator-এর গুণ বিচার করতে হয়। Unbiasedness ছিল একটা কায়দা। আরেকভাবেও আমরা কাজটা করতে পারি। যত n বাড়বে ততই যদি $\widehat{\theta}_n$-টা θ-র কাছে এগিয়ে যায়, সেটাকে বলব consistency. সংজ্ঞাটা দিলাম নীচের অংকের উত্তরে।

Example 3: Define a consistent estimate of a parameter of a population distribution.[2]

(2004,2005,2007,2009,2012)

SOLUTION:

> **DEFINITION: Consistent**
>
> Let $X_1, ..., X_n$ be a random sample from a distribution belonging to a known family. Let θ be an unknown parameter, and let $\widehat{\theta}_n(X_1, ..., X_n)$ be an estimator of θ. We say that $\widehat{\theta}_n(X_1, ..., X_n)$ is consistent for θ if for all distributions in the family
>
> $$\widehat{\theta}_n(X_1, ..., X_n) \xrightarrow{P} \theta \text{ as } n \to \infty.$$

■

এবার এই ধারণা দুটোর কিছু প্রয়োগ দেখব।

33.1.1 Sample mean এবং sample variance

প্রথমে যে প্রয়োগগুলো দেখতে চলেছি তাদেরকে গোড়াতেই একটা theorem হিসেবে লিখে নিই, কারণ এরা গুরুত্বপূর্ণ জিনিস।

> **THEOREM**
>
> Let $X_1, ..., X_n$ be IID with mean μ and variance $\sigma^2 < \infty$. Then
>
> 1. $\overline{X}$ is an unbiased, consistent estimator of μ.
> 2. $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is a biased estimator of σ^2.
> 3. $\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is an unbiased estimator of σ^2.

এই theorem-টার বিভিন্ন অংশের প্রমাণ নিয়েই নীচের অংক কয়টা।

Example 4: Prove that sample mean is always unbiased estimate of the population mean. **(2004,2005,2007,2008,2012)**

SOLUTION:

> Let $X_1, ..., X_n$ be a random sample from an (infinite) population with mean μ.

Family-টা বলে নেওয়া ভালো--

> Here we are considering the family of all distributions with finite mean.
>
> Shall show that $\overline{X}_n = \frac{1}{n}\sum_1^n X_i$ is an unbiased estimator of μ, ie, for all distributions in the family
>
> $$E(\bar{X}_n) = \mu.$$
>
> Now
>
> $$\begin{aligned} E(\bar{X}_n) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} \times n\mu = \mu, \end{aligned}$$
>
> as required.

■

Example 5: Prove that sample mean is always consistent estimate of the population mean. **(2005,2007,2008,2012)**

SOLUTION:

> Let $X_1, ..., X_n$ be a random sample from some distribution with mean μ.

এবার family-টা বলে নেওয়া যাক--

> Here we are considering the family of all distributions with finite mean and variance.

লক্ষ কর আমরা variance-টাও finite রেখেছি, যাতে Chebyshev inequality লাগানো যায়। Variance-টা finite হলে অবশ্য mean-ও finite হতে বাধ্য। সুতরাং ওটা আলাদা করে না বললেও চলত।

> Shall show that $\bar{X}_n$ is a consistent estimator of μ, ie, for all distributions in the family $\overline{X}_n \xrightarrow{P} \mu$,

এটা আসলে WLLN ছাড়া আর কিছুই নয়। কিন্তু স্রেফ WLLN বলে ছেড়ে দেওয়াটা খারাপ দেখায়, তাই WLLN-এর প্রমাণটা এখানে ঢুকিয়ে দিই--

> ie,
>
> $$\forall \epsilon > 0 \quad \lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0.$$
>
> Take any $\epsilon > 0$.
>
> Then by Chebyshev inequality
>
> $$P(|\bar{X}_n - E(\bar{X}_n)| > \epsilon) \leq \frac{var(\bar{X}_n)}{\epsilon^2}.$$
>
> Now $E(\bar{X}_n) = \mu$, and
>
> $$var(\bar{X}_n) = var\left(\frac{1}{n}\sum_1^n X_i\right) = \frac{1}{n^2}\sum_1^n \sigma^2 = \frac{\sigma^2}{n}.$$
>
> So
>
> $$P(|\bar{X}_n - E(\bar{X}_n)| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2 n} \to 0,$$
>
> as $n \to \infty$, as required.

■

Example 6: Prove that sample variance is not an unbiased estimator of the population variance.[4] **(2004,2005,2013)**

SOLUTION:

> Let $X_1, ..., X_n$ be a random sample from some distribution with variance σ^2.

এবার family-টা বলে নেওয়া যাক--

> Here we are considering the family of all distributions with finite variance.
>
> Shall show that sample variance is not an unbiased estimator of population variance, ie, for some distribution in the family
>
> $$E(S_n^2) \neq \sigma^2,$$
>
> where
>
> $$S_n^2 = \frac{1}{n}\sum_1^n X_i^2 - (\bar{X}_n)^2.$$

লক্ষ কর এখানে আমরা S_n^2-কে $\frac{1}{n}\sum(X_i - \overline{X})^2$ আকারে লিখিনি। সেভাবে লিখলে অংকটা একটু বেশী জটিল হত।

> Now $E(X_i^2) = \text{var}(X_i) + (E(X_i))^2 = \sigma^2 + \mu^2.$

এখানে আমরা এই তথ্যটা ব্যবহার করেছি যে, $\text{var}(X) = E(X^2) - \big(E(X)\big)^2$.

> Also $E(\bar{X}_n^2) = \text{var}(\bar{X}_n) + (E(\bar{X}_n))^2 = \frac{\sigma^2}{n} + \mu^2.$

এখানে আমরা ব্যবহার করেছি $\text{var}(\overline{X}_n) = \frac{\sigma^2}{n}$. এর প্রমাণটা ছোট্টো করে ঢুকিয়ে দিই--

> ⟦ Because:
>
> $\text{var}(\overline{X}_n) = \text{var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n \text{var}(X_i) = \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{\sigma^2}{n}$, since X_i's are IID.
>
> ⟧
>
> So
>
> $$\begin{aligned} E(S_n^2) &= E\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}_n^2\right) \\ &= \frac{1}{n}\sum_{i=1}^n E(X_i^2) - E(\bar{X}_n^2) \\ &= \frac{1}{n}\sum_{i=1}^n (\sigma^2 + \mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right) \\ &= \cdots \\ &= \frac{n-1}{n}\sigma^2 \neq \sigma^2, \end{aligned}$$

for any $\sigma^2 > 0$, as required.

■

Exercise 1: Prove that sample mean is an UE for population mean but sample variance is not.[4] **(2010)**
HINT: এটা মোটেই নতুন কোনো অংক নয়। এখানে UE মানে হল unbiased estimator. ■

Example 7: Prove that sample mean is an unbiased estimate of the population mean, but the sample variance is not an unbiased estimate of the population variance. Obtain an unbiased estimate of the population variance. **(2004.8a)**
SOLUTION: প্রথম দুটো অংশ তো আগের অংকগুলোতেই হয়েছে। এবার শেষ অংশটা করি--

Last part:

We have seen that $E(S_n^2) = \frac{n-1}{n}\sigma^2$.
So
$$E\left(\frac{n}{n-1}S_n^2\right) = \frac{n}{n-1}E(S_n^2) = \frac{n}{n-1} \times \frac{n-1}{n}\sigma^2 = \sigma^2.$$
Hence the statistic
$$T = \tfrac{n}{n-1}S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$
is an unbiased estimator of σ^2.

■

Exercise 2: A random sample $(x_1, x_2, ..., x_n)$ is drawn from a normal(m, s) population. Find the maximum likelihood estimate of parameters m, s. Hence deduce that maximum likelihood estimate of the mean is consistent.[6] **(2004.8b)**
HINT:
MLE দুটো তো দ্বিতীয় অধ্যায়েই বার করেছিলাম--
$$\widehat{m} = \overline{X} \text{ এবং } \widehat{s} = \sqrt{\frac{1}{n}\sum_{1}^{n}(X_i - \overline{X})^2}.$$
এর মধ্যে $\widehat{m} = \overline{X}$ যে consistent, সে তো প্রমাণ করলামই। ■

33.1.2 WLLN দিয়ে consistency প্রমাণ

Consistency দেখানো মানে $\widehat{\theta}_n \xrightarrow{P} \theta$ দেখানো। Convergence in probability-ওয়ালা একটাই theorem আমরা শিখেছি, সেটা হল WLLN. অনেক সময়েই consistency দেখানোর কাজে WLLN খুব সাহায্য করে। WLLN-এর বক্তব্যটা একটু মনে করে নিই--

যদি একটা sequence থাকে $\{X_i\}_i$, যেখানে X_i-রা হল IID, তাহলে--

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} E(X_1).$$

এর জন্য অবশ্যই expectation-টার অস্তিত্ব থাকতে হবে, নইলে আর $E(X_1)$ লিখছি কী করে! কিন্তু আমরা যেভাবে WLLN প্রমাণ করেছিলাম, তার জন্য আরেকটু জোরালো শর্ত চাই। খালি expectation-এর অস্তিত্ব থাকলেই চলবে না, $\text{var}(X_1)$-কেও finite হতে হবে[2]।

এখানে খালি X_1-এর expectation আর variance লিখেছি দেখে অবাক হয়ো না। মনে রেখো যে, X_i-দের সবারই distribution একই হওয়ায় সব X_i-দেরই একই expectation আর variance.
WLLN-এর অধিকাংশ প্রয়োগের সময়েই আমরা X_i-দের কোনো function নিয়ে কাজ করি। ধরো $h(x)$ কোনো একটা function. তাহলে $h(X_i)$-রাও সবাই IIDহবে (যেহেতু X_i-রা IID ছিলই, আর সবার উপরেই একই function লাগানো হয়েছে)। যদি $\text{var}(h(X_1)) < \infty$ হয় অমনি WLLN থেকে বলা যাবে--

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \xrightarrow{P} E(h(X_1)).$$

মনে রেখো যে, $\text{var}(h(X_1)) < \infty$ হওয়ায় $E(h(X_1))$-ও finite হতে বাধ্য।
এই ধারণাটা কাজে লাগিয়ে বিভিন্ন ক্ষেত্রে consistency দেখানো সম্ভব। এবার সেরকম কয়েকটা উদাহরণ দেখব।

Example 8: Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from a normal population with mean zero and standard deviation σ. Show that $\frac{\sum_{i=1}^{n} X_i^2}{n}$ is an unbiased estimate of σ^2. Examine whether this estimate is consistent.[6] **(2014.5b)**
SOLUTION: আগের অংকের প্রয়োগ। ■

Exercise 3: Find the maximum likelihood estimate of σ^2 of a normal(m, σ) population if m is known, and show that the estimate is unbiased and consistent.[4+1+2] **(2012.4c)**
HINT:
প্রথম অংশ তো আগেই করেছি (দ্বিতীয় অধ্যায়ে)। MLE-টা হয়েছিল $\frac{1}{n}\sum_1^n (X_i - m)^2$. এটার expectation দেখাতে হবে σ^2. সেটা খুবই সহজ। একটু ধরিয়ে দিই। $E\big((X_i - m)^2\big) = \text{var}(X_i - m) + \big(E(X_i - m)\big)^2$.
Consistency আসবে একেবারেই আগের অংকটার মত করে।
■

Example 9: A population is defined by the density function

$$f(x, \alpha) = \frac{x^{\ell-1} e^{x/\alpha}}{\alpha^{\ell}\Gamma(\ell)}, \quad 0 < x < \infty,$$

ℓ being a known constant. Estimate the parameter α by the method of maximum likelihood, and show that the estimate is consistent and unbiased.[7] **(2010.4b)**
SOLUTION: এই অংকের প্রথম অংশটা (মানে α-র MLE বার করা অংশটা) দেওয়া ছিল দ্বিতীয় অধ্যায়ে। উত্তরটা এসেছিল $\widehat{\alpha} = \frac{\sum_1^n X_i}{n\ell}$. এবার দেখাব যে, সেটা consistent এবং unbiased.
এর জন্য X_i-দের expectation বার করতে হবে।

[2]Variance যদি finite নাও হয়, তাও WLLN প্রমাণ করা যায়। সেই প্রমাণটা একটু জটিল, তাই এই বইতে সে প্রসঙ্গে যাব না।

Here $\ell > 0$ for $f(x, \alpha)$ to a PDF.

Unbiased:

$$E(X_i) = \int_{-\infty}^{\infty} x f(x, \alpha) dx = \int_0^{\infty} \frac{x^\ell e^{x/\alpha}}{\alpha^\ell \Gamma(\ell)} dx = \cdots = \alpha\ell.$$

এটা এল Gamma integral লাগিয়ে।

$$E(\widehat{\alpha}) = E\left(\frac{\sum_{i=1}^n X_i}{n\ell}\right) = \frac{\sum_{i=1}^n E(X_i)}{n\ell} = \frac{\sum_{i=1}^n \alpha\ell}{n\ell} = \frac{n\alpha\ell}{n\ell} = \alpha,$$

as required.

Consistency দেখানোর জন্য $\text{var}(X_i) < \infty$ দেখাতে হবে। যেহেতু $E(X_i)$ বার করেই ফেলেছি, তাই এর জন্য $E(X_i^2) < \infty$ দেখানোই যথেষ্ট।

Consistent:

$$E(X_i^2) = \int_{-\infty}^{\infty} x^2 f(x, \alpha) dx = \int_0^{\infty} \frac{x^{\ell+1} e^{x/\alpha}}{\alpha^\ell \Gamma(\ell)} dx < \infty,$$

by property of gamma integrals.

So $\text{var}(X_i) < \infty$.

Hence by WLLN we have

$$\frac{1}{n} \sum_1^n X_i \xrightarrow{P} \alpha\ell,$$

$\therefore \widehat{\alpha} = \frac{\sum_1^n X_i}{n\ell} \xrightarrow{P} \alpha$ when $n \to \infty$, as required.

■

33.1.3 Chebyshev inequality দিয়ে consistency প্রমাণ

আমরা WLLN দিয়ে consistency প্রমাণের অনেকগুলো উদাহরণ দেখলাম। WLLN-এর প্রমাণটা যদি তোমার মনে থেকে থাকে, তবেই দেখবে যে WLLN-এর মূল শক্তিটা আসে Chebyshev inequality থেকে। সুতরাং Chebyshev inequality দিয়েও সরাসরি consistency প্রমাণ করা যায়। সেটা নিয়েই নীচের অংকটা।

Example 10: For a sample of size n from any population, let θ be an unbiased estimate of population parameter α and $\text{var}(\theta) \to 0$ as $n \to \infty$. Prove that θ is a consistent estimate of α.[3] **(2009.4c)**

SOLUTION: এই অংকে ব্যবহৃত notation একেবারেই standard নয়। ছোটো হাতের গ্রীক অক্ষর সাধারণতঃ parameter বোঝাতে ব্যবহৃত হয়। এখানে θ দিয়ে বুঝিয়েছে estimator.

To show

$$\forall \epsilon > 0 \quad \lim_{n\to\infty} P(|\theta - \alpha| \geq \epsilon) = 0.$$

Take any $\epsilon > 0$.

By Chebyshev inequality

$$P(|\theta - E(\theta)| \geq \epsilon) \leq \frac{\text{var}(\theta)}{\epsilon^2}.$$

Since θ is unbiased for α, so $E(\theta) = \alpha$.

Hence

$$P(|\theta - E(\theta)| \geq \epsilon) = P(|\theta - \alpha| \geq \epsilon) \leq \frac{\text{var}(\theta)}{\epsilon^2} \to 0,$$

since $\text{var}(\theta) \to 0$ as $n \to \infty$.

Since any probability ≥ 0, hence by sandwich law

$$P(|\theta - E(\theta)| \geq \epsilon) \to 0 \text{ as } n \to \infty,$$

as required.

■

33.2 Minimum variance unbiased estimator

কোনো estimator-এর গুণ বিচারের দুটো কায়দা শিখেছি আমরা--unbiasedness আর consistency. এছাড়া অন্যান্য বিভিন্ন কায়দাও আছে। একটা হল estimator-টার variance কত দেখা (বা তার square root, মানে standard error, দেখা)। এটা খুব বেশী হওয়া ভালো কথা নয়। বেশী variance (বা standard error)-ওয়ালা estimator ব্যবহার করে estimate করতে যাওয়া আর কম্পমান হাতে বন্দুক ছোঁড়া কতকটা একইরকম ব্যাপার। স্বাভাবিকভাবেই আমরা চাইব যেন বন্দুকটা ঠিক দিকে তাক করা থাকে এবং হাতও যথাসম্ভব স্থির থাকে। Estimation-এর ভাষায় আমরা চাইছি যেন আমাদের estimator-টা unbiased হয় এবং তার variance-ও যথাসম্ভব কম হয়। এই দুই গুণ মিলিয়ে যা দাঁড়ায় তার একটা নাম আছে--Minimum variance unbiased estimator. বিশদ সংজ্ঞা চেয়েছে নীচের অংকে।

Example 11: What is an MVUE?[1] **(2010.5c)**

SOLUTION:

DEFINITION: **MVUE**

By a minimum variance unbiased estimator (MVUE) of a parameter θ based on a random sample, we mean an unbiased estimator of θ that has the minimum possible variance among all unbiased estimators of θ based on that sample.

ভালো করে খুঁটিয়ে পারলে লক্ষ করবে যে, এখানে minimum-টা নেওয়া হচ্ছে যাবতীয় unbiased estimator-দের মধ্যে,

যাবতীয় estimator-দের মধ্যে নয়। ■

DAY 34
Confidence interval

Estimation কী জিনিস সেটা আমরা আগেই শিখেছি। একটা parameter যদি থাকে θ, তবে আমরা তাকে একটা সংখ্যা দিয়ে estimate করছিলাম। যেমন একটা কয়েন টস করা হল 10-বার, তার মধ্যে head পড়ল ঠিক 4-বার। তার ভিত্তিতে $P(\text{head})$-কে estimate করলাম $\frac{4}{10}$. এর মানে অবশ্যই এই নয় যে, $P(\text{head})$ একদম $\frac{4}{10}$ হতেই বাধ্য। কিন্তু মোটামুটি তার ধারেকাছে কিছু একটা হবে, এমনটা আশা করা অসঙ্গত নয়। এরকম ক্ষেত্রে স্বভাবতঃই প্রশ্ন ওঠে যে, "ধারেকাছে" বলতে ঠিক কতটা কাছে বোঝানো হচ্ছে? সেইটা বোঝানোর একটা কায়দা হল খালি $\frac{4}{10}$ না লিখে তার সঙ্গে $\pm$ কিছু একটা জুড়ে দেওয়া, যেমন $\frac{4}{10} \pm 0.1$, অর্থাৎ 0.3 থেকে 0.5-এর মধ্যে কিছু একটা। যদি head-এর সংখ্যাটা X বলি তবে আমরা এই interval-টা পাচ্ছি, $\left(\frac{X}{10} - 0.1, \frac{X}{10} + 0.1\right)$. একে বলব একটা confidence interval. অবশ্য এখানেও আমরা একেবারে নিশ্চিত হতে পারছি না যে, $P(\text{head})$ সত্যই এর মধ্যে পড়বেই, কিন্তু পড়ার probability-টা অন্ততঃ আগের চেয়ে বেশী। যদি তাতেও মন না ওঠে তবে আমরা 0.1-এর জায়গায় 0.2 বা আরও বড় কিছু নিয়ে interval-টার বিস্তার বাড়াতে পারি, যাতে probability-টা যথেষ্ট বড় হয়। অবশ্য এই probability-টা আবার কয়েনটার $P(\text{head})$-র উপর নির্ভর করবে। যদি দেখা যায় যে, এই এটা সবসময়েই (মানে যেকোনো $P(\text{head})$-এর জন্যই) $\geq \beta$, তবে আমরা বলব যে এটা একটা β **level confidence interval.** সাধারণতঃ β-কে $1-\alpha$ আকারে লেখা হয়, অর্থাৎ α হল ভুল করার probability (মানে parameter-টা interval-টার বাইরে পড়ার probability)।

দ্বিতীয় অধ্যায়ে আমরা যখন estimation শিখেছিলাম (MLE ইত্যাদি) তখন আমরা কোনো parameter-কে একটা সংখ্যা দিয়ে estimate করছিলাম। তাকে বলে **point estimation.** ব্যাপারটা যেন কতকটা বল্লম দিয়ে মাছ শিকারের মত। যে parameter-টা estimate করছি সেটা যেন মাছটা, জলের নীচে আছে বলে তাকে চোখে দেখা যাচ্ছে না। খালি তার জলের উপরে একটা আলোড়ন টের পাওয়া যাচ্ছে। সেই আলোড়নটা হল data. তার ভিত্তিতে আমরা বল্লম ছুঁড়ছি (মানে estimate করছি)। এখানে estimate-টা হল একটা সংখ্যা, সেটা যেন বল্লমের ডগাটা। সেই ডগাটা যদি একেবারে লক্ষ্যভেদ করতে পারে, তবে ভালো, নয়তো সামান্য এদিক ওদিক হলেই

Confidence interval ব্যাপারটাও সেইরকম, খালি এখানে একটা সংখ্যার বদলে একটা interval. ঠিক যেন বল্লমের বদলে জাল ব্যবহার করা হচ্ছে। ফলে মাছটা ধরা পড়ার সম্ভাবনা বেশী।

Example 12: What is meant by confidence interval for a parameter of a distribution?[1] **(2004,2006)**

SOLUTION:

DEFINITION: **Confidence interval**

Let $X_1, ..., X_n$ be IID data from a distribution belonging to known family with an unknown parameter θ. Then by a confidence interval of confidence level $1-\alpha$ we mean an interval I (depending on the data) such that for all distributions in the family

$$P(\theta \in I) = 1 - \alpha.$$

■

এখানে একটা লেখার কৌশল করেছি, interval-টাকে (a, b) বা $[a, b]$ ইত্যাদি না লিখে I নাম দিয়েছি, তাতে $(a, b), [a, b], [a, b), (a, \infty), (-\infty, b]$ ইত্যাদি সবকিছুই একসঙ্গে লেখা হয়ে গেল।

Example 13: Define confidence interval. [1] **(2012.5a)**

SOLUTION: আগের অংকটাই। ■

এবার আমরা confidence interval বার করার কিছু উদাহরণ দেখব।

34.1 Normal mean

Example 14: Find the confidence interval for the parameter m of a normal(m, σ) population on the basis of a sample of size n having confidence coefficient $(1-\alpha)$. Assume that σ is known. $(0 < \alpha < 1)$.[5] **(2010,2012)**

SOLUTION: প্রশ্নের ভাষার দোষে সবকিছু কেমন যেন তালগোল পাকিয়ে গেছে। আগে জটটা ছাড়িয়ে নিই। একটা sample আছে যার সাইজ n. তার মানে $X_1, ..., X_n$ যারা IID. এদের distribution আছে $N(m, \sigma)$ নামক family-তে যেখানে σ একটা জানা সংখ্যা (অর্থাৎ প্রশ্নটার উত্তরে σ ব্যবহার করতে পারি)। কিন্তু m অজানা। এই m-এর জন্যই একটা confidence interval বার করতে হবে, যার level হওয়া চাই $1 - \alpha$. এখানে যে confidence coefficient কথাটা বলেছে, সেটা আসলে confidence level-এরই সমার্থক।

> Let $X_1, ..., X_n$ be IID $N(m, \sigma)$.
> Then we know that $\bar{X} \sim N\left(m, \frac{\sigma}{\sqrt{n}}\right)$, and hence
> $$\frac{\bar{X} - m}{\sigma/\sqrt{n}} \sim N(0, 1).$$

এইটা আমরা অষ্টম অধ্যায়ে শিখেছিলাম। এখানে এটা কী কাজে লাগবে? যাতে এমন একটা কিছু জিনিস পাই যেটা data এবং আমাদের parameter-টা দিয়ে তৈরী, এবং যার distribution-এর মধ্যে কোনো অজানা কিছু নেই। এরকম জিনিসের একটা নাম আছে, এদের বলে **pivot.** মানে এখানে $\frac{\bar{X}-m}{\sigma/\sqrt{n}}$ হল একটা pivot.

একবার একটা pivot পেয়ে গেলেই কেল্লা প্রায় ফতে। খালি একটাই কাজ বাকি--এমন একটা interval নেওয়া, যাতে pivot-টা ওর মধ্যে থাকার probability হয় $1 - \alpha$. এই কাজটা করা যাবে, কারণ pivot-এর distribution আমরা সম্পূর্ণভাবে জানি। যেমন এখানে distribution-টা হল $N(0, 1)$. সুতরাং এখানে এমন একটা interval নিতে হবে যাতে Fig 1-এর মত area-টা হয় $1 - \alpha$. এই কাজটা নানাভাবে করা যায়। একটা কায়দা হল Fig 2-এর মত 0-র দুপাশে

Fig 1

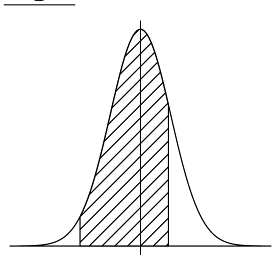
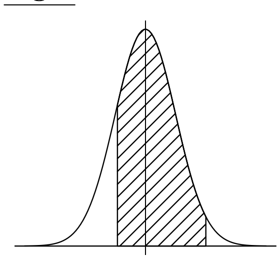


Fig 2



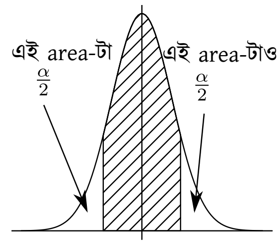


Fig 3

symmetric-ভাবে নেওয়া, মানে $a = -c$আর $b = c$ নেওয়া। এরকম c-কে ফর্মুলা দিয়ে লেখা সহজ। এখানে $-c$ থেকে c পর্যন্ত মোট area হল $1-\alpha$. সুতরাং এর বাইরের area-টা হবে α, যেটা symmetry-র জন্য সমান দুইভাগে ভাগ হয়ে আছে, একভাগ আছে $-c$-এর বাঁদিকে, অন্যভাগটা c-এর ডানদিকে। তাই c-এর ডানদিকে মোট area হল $\frac{\alpha}{2}$. অতএব c-র বাঁদিকের মোট area হবে $1-\frac{\alpha}{2}$. এবার মনে রেখো যে কোনো x-এর বাঁদিকের মোট area-টাকে বলে $\Phi(x)$, যেটা $N(0,1)$-এর CDF. তাই $\Phi(c) = 1-\frac{\alpha}{2}$. সুতরাং--

> Let $c = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$.
> Then
> $$P\left(\frac{\bar{X}-m}{\sigma/\sqrt{n}} \in (-c,c)\right) = 1-\alpha,$$
> or
> $$P\left(m \in \left(\bar{X}-\tfrac{c\sigma}{\sqrt{n}}, \bar{X}+\tfrac{c\sigma}{\sqrt{n}}\right)\right) = 1-\alpha.$$

এইটা একবারে কী করে ধাঁ করে লিখে দিলাম, বুঝলে? Interval-টার দুই প্রান্তকেই প্রথমে $\frac{\sigma}{\sqrt{n}}$ দিয়ে গুণ করে সেটাকে $\overline{X}$ থেকে বিয়োগ করেছি।

> So one possible $1-\alpha$ level confidence interval is
> $$\left(\bar{X}-\tfrac{c\sigma}{\sqrt{n}}, \bar{X}+\tfrac{c\sigma}{\sqrt{n}}\right),$$

■

মনে রেখো যে এটাই একমাত্র উত্তর নয়। চাইলে আমরা interval-টাকে $(-\infty, b)$-এর মতও নিতে পারতাম (Fig 3)। এখানে $b = \Phi^{-1}(1-\alpha)$ নিলেই কাজ হবে।

Example 15: What do you mean by confidence interval for mean of normal(m,σ) population when (i) σ is known and (ii) σ is unknown.[6] **(2008.5b)**

SOLUTION:

> Let $X_1, ..., X_n$ be IID $N(m,\sigma)$.
> (i) When σ is known, then a confidence interval of level $1-\alpha$ for m means a random interval I, which is a function of the sample, such that for all distributions in the given family
> $$P(m \in I) = 1-\alpha.$$

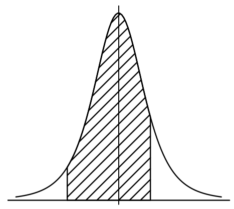


Fig 4

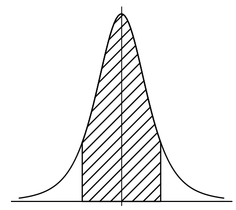


Fig 5

> One such interval may be obtained as follows.
> We know that
> $$\frac{\bar{X}-m}{\sigma/\sqrt{n}} \sim N(0,1).$$
> Let $c = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$.
> Then
> $$I = \left(\bar{X} - \tfrac{\sigma}{\sqrt{n}}c, \bar{X} + \tfrac{\sigma}{\sqrt{n}}c\right)$$
> is one possible $(1-\alpha)$-level confidence interval for m.

এইটা বস্তুতঃ আগের অংকটাই ছিল। এবার দেখি σ জানা না থাকলে কী হত। এখানেও মূল কায়দাটা একই, প্রথমে একটা pivot বার করতে হবে।

> (ii) When σ is unknown, then a confidence interval of level $1-\alpha$ for m means a random interval I, which is a function of the sample, such that for all distributions in given family
> $$P(m \in I) = 1 - \alpha.$$
> From standard result we know that
> $$\frac{\bar{X}-m}{S/\sqrt{n}} \sim t_{(n-1)},$$
> where
> $$S = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}}.$$

ব্যস্, এটাই আমাদের pivot. এবার এমন একটা interval নেব যাতে Fig 4-এর মত area-টা $1-\alpha$ হয়। একটা কায়দা হল 0-র দুপাশে symmetric-ভাবে নেওয়া, মানে Fig 5-এর মত করে।

> Let $c = F^{-1}\left(1-\frac{\alpha}{2}\right)$, where $F(x)$ is the distribution function of the $t_{(n-1)}$ distribution. Then
> $$I = \left(\bar{X} - \tfrac{S}{\sqrt{n}}c, \bar{X} + \tfrac{S}{\sqrt{n}}c\right)$$

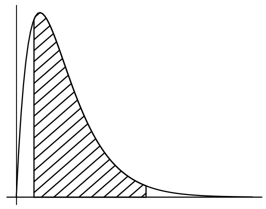


Fig 6

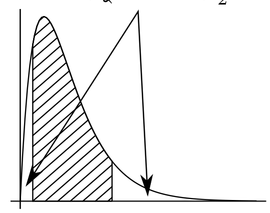


Fig 7

is one possible $(1-\alpha)$ level confidence interval for m.

■

34.2 Normal variance

এবার আমরা আরও কিছু উদাহরণ দেখব যেখানে আমরা normal distribution-এর variance-এর জন্য confidence interval বার করব।

Example 16: Find a confidence interval for the parameters of a normal(m, σ) population with confidence coefficient $1-\alpha$ on the basis of a random sample of size n drawn from the population.[6] **(2006.5a)**

SOLUTION: এই অংকটার ঠিক মানে হয় না। যদি m আর σ দুটোই অজানা হয়, তবে তাদের জন্য একটা interval দিয়ে কাজ চলতে পারে না, কারণ interval হল $\mathbb{R}$-এর subset. আমাদের দরকার $\mathbb{R}^2$-এর কোনো subset. সেরকম subset-কে বলে **confidence region.** সেটা বার করারও কায়দা আছে বটে, কিন্তু সেটা এই বইয়ের পাল্লার বাইরে। সম্ভবতঃ এই অংকে আসলে বলতে চাইছে যে, m-এর জন্য একটা confidence interval আর σ-র জন্য আলাদা করে আরেকটা confidence interval বার করতে।

m-এর জন্য কী করে এগোবে সেটা তো আগের অংকেই বলেছি। σ-র জন্যে কী করবে বলি।

We know that

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 \sim \chi^2_{(n-1)}.$$

এটা আমরা আগের অধ্যায়েই শিখেছি। লক্ষ কর যে, এই distribution-টার মধ্যে কোনো অজানা জিনিস নেই, সুতরাং $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2$ হল একটা pivot. এবার আমাদের দরকার এমন একটা interval যাতে Fig 6-র মত area-টা হয় $1-\alpha$. এখানে যেহেতু distribution-টা মোটেই 0-র দুপাশে symmetric নয়, তাই symmetric-ভাবে interval-টা নেওয়ার কোনো কারণ নেই। একটা কায়দা হতে পারে Fig 7-এর মত করে নেওয়া।

Let $c = F^{-1}(\alpha)$, where $F(x)$ is the distribution function of $\chi^2_{(n-1)}$ distribution.

Then

$$P\left(\frac{\sum(X_i-\bar{X})^2}{\sigma^2} > c\right) = 1-\alpha,$$

and hence

$$P\left(\frac{\sigma^2}{\sum(X_i-\bar{X})^2}<\frac{1}{c}\right)=1-\alpha,$$

or

$$P\left(\sigma^2<\frac{\sum(X_i-\bar{X})^2}{c}\right)=1-\alpha.$$

Since $\sigma^2>0,$ hence one possible confidence interval is

$$\left(0,\frac{\sum(X_i-\bar{X})^2}{c}\right).$$

■

34.3 Proportion

লক্ষ কর যে, confidence interval বার করার মূল কায়দাটা হল একটা pivot বার করতে পারা। এরকম pivot সবসময়ে পাওয়া সহজ নাও হতে পারে, এমন কি অসম্ভবও হতে পারে। তখন অনেক সময়ে আমাদেরকে কোনো approximate pivot দিয়ে কাজ চালাতে হয়, অর্থাৎ এমন কোনো জিনিস যেটা--

- খালি দুটো জিনিসের উপর নির্ভর করে--
  - data-র উপর,
  - যে parameter-এর confidence interval বার করছি, তার উপর,
- এবং যার distribution পুরো জানা না থাকলেও, কোনো একটা জানা distribution-এর খুব কাছাকাছি হয়।

নীচের অংকে একটা উদাহরণ আছে, দেখলে ব্যাপারটা বুঝতে সুবিধা হবে।

Example 17: Find an approximate confidence interval with a given confidence coefficient $1-\epsilon$, $(0<\epsilon<1)$ for p of a binomial(n,p) population assuming that n is known and large.[7] **(2004,2005)**

SOLUTION:

Since n is large, we can apply CLT to assume that

$$T=\frac{X-np}{\sqrt{np(1-p)}}$$

Fig 8

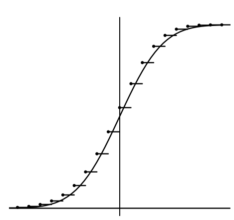
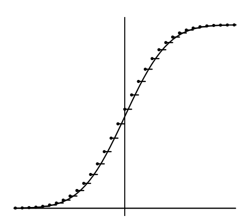


Fig 9



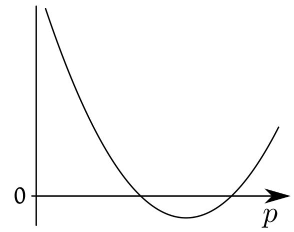


Fig 10

has an asymptotic $N(0,1)$ distribution.

সুতরাং এই জিনিসটা হল আমাদের approximate pivot. এর আসল distribution-টা কিন্তু বেজায় বদখত কিছু একটা, এবং সেটা আমাদের অজানা p-এর উপরেও নির্ভর করে। কিন্তু বদখত হলে কি হয়, distribution-টা $N(0,1)$-এর খুবই কাছে। Fig 8 আর Fig 9 দেখলে ব্যাপারটা স্পষ্ট হবে। Fig 8-এ যে টানা লাইনটা রয়েছে সেটা হল $N(0,1)$-এর CDF, আর step function-টা হল T-এর CDF,যেখানে $n = 100$ আর $p = 0.5$ নিয়েছি। এইটাই হল সেই বদখত CDF, যেটাকে আমরা এই ছবিটা আঁকার জন্য কম্পিউটার দিয়ে বার করেছি। লক্ষ করো যে এই CDF-টা $N(0,1)$-এর CDF-এর বেশ কাছেই আছে। Fig 9-এর ছবিটা ঠিক একই জিনিস, খালি এবার $p = 0.1$ নিয়েছি (n একই আছে)। লক্ষ কর যে, step function-টা সামান্য বদলেছে, যদিও এটাও $N(0,1)$-এর CDF-এর কাছেই রয়েছে।

Let $c = \Phi^{-1}\left(1 - \frac{\epsilon}{2}\right)$.
Then $P(T \in (-c,c)) \approx 1-\epsilon$.

এইবার আমরা $T \in (-c,c)$-কে দলাইমলাই করে p দিয়ে লিখব। T-এর নীচের তলায় একটা square root আছে। সেটাকে তাড়াবার জন্য square করে নিলে সুবিধা হবে--

Now

$$
\begin{aligned}
T \in (-c,c) \quad &\iff \quad T^2 < c^2 \\
&\iff \quad \frac{(X-np)^2}{np(1-p)} < c^2 \\
&\iff \quad (X-np)^2 < c^2 np(1-p) \quad [\because 0 < p < 1] \\
&\iff \quad n^2p^2 - 2nXp + X^2 < -c^2p^2 + c^2p \\
&\iff \quad (n^2+c^2)p^2 - (2nX+c^2)p + X^2 < 0.
\end{aligned}
$$

জিনিসটা দেখতে হয়তো খুব উপাদেয় লাগছে না। কিন্তু একটু গ্রাফ দিয়ে ভাবলেই বুঝবে যে ব্যাপারটা তেমন জটিল কিছু নয়। বাঁদিকে আছে p-এর একটা quadratic. যার গ্রাফ হল Fig 10-এর মত একটা parabola. আমাদের দেখতে হবে p-এর কোন কোন value-র জন্য সেটা negative হয়, মানে x-axis-এর নীচে থাকে। উত্তরটা তো বুঝতেই পারছ ছবি দেখে--quadratic-টার দুটো root-এর মাঝখানে সব value-র জন্য। অবশ্য একটা প্রশ্ন হল সত্যিই quadratic-টার দুটো real root আছে তো? সে প্রশ্নের নিরসন হবে discriminant-টা বার করলেই--

The discriminant of the LHS is

$$\Delta = (2nX+c^2)^2 - 4(n^2+c^2)X^2 = \cdots = c^4 + 4c^2X(n-X) > 0,$$

since $X \in \{0,1,...,n\}$.

Thus the quadratic has distinct real roots

$$p_1 = \frac{-(2nX + c^2) - \sqrt{\Delta}}{2(n^2 + c^2)} \text{ and } p_2 = \frac{-(2nX + c^2) + \sqrt{\Delta}}{2(n^2 + c^2)}.$$

So $T \in (-c, c) \iff p \in (p_1, p_2)$, which is an approximate $(1-\epsilon)$-level confidence interval.

■

হ্যাঁ, মানছি যে, ধাপগুলো বেশ বিচ্ছিরি দেখতে। দোষটা অবশ্য আমার নয়, দোষ হল T-এর নীচের তলার ওই square root-টার। স্বভাবতঃই প্রশ্ন ওঠে--অন্য কোনো approximate pivot আছে কিনা, যেটা দিয়ে ভদ্রমত একটা confidence interval পাওয়া যায়? উত্তর হল--হ্যাঁ, পাওয়া যায়। তার জন্যে লক্ষ কর যে, $\frac{X}{n} \xrightarrow{P} p$, মানে n যথেষ্ট বড় হলে $\frac{X}{n} \approx p$ হবে। ধরো $\frac{X}{n}$-কে $\widehat{p}$ নাম দিলাম। তাহলে $\widehat{p}$ কিন্তু আমাদের সম্পূর্ণ জানা জিনিস, কারণ n আর X দুটোই আমরা জানি। আমাদের T ছিল এইরকম--

$$T = \frac{X - np}{\sqrt{np(1-p)}}.$$

এর নীচের তলার square root-এর মধ্যে p থাকাতেই সমস্যা হচ্ছিল, সেই জায়গায় $\widehat{p}$ বসিয়ে দিলে পাব

$$\widehat{T} = \frac{X - np}{\sqrt{n\widehat{p}(1-\widehat{p})}}.$$

লক্ষ কর যে, উপর তলার p-টার গায় কিন্তু হাত দিই নি, কারণ যদি আমাদের pivot থেকে সব p-ই বাদ চলে যায়, তবে আর সেই pivot দিয়ে p-এর confidence interval বেরোবে কী করে?
এখন আমাদের আশা হল, যেহেতু $\widehat{p} \approx p$, তাই নিশ্চয়ই $\widehat{T}$-ও T-এর মতই asymptotically $N(0,1)$ হবে। তা যে সত্যিই হবে, সেটা অংক কষে প্রমাণ করা যায়, কিন্তু তার জন্য Slutsky's theorem বলে একটা জিনিস লাগে যেটা আমাদের আলোচনার পাল্লার বাইরে। সেই কারণেই আমাদের সমাধানে আমরা T নিয়ে কাজ করেছিলাম। যাই হোক, $\widehat{T}$ নিয়ে কাজ করলে উত্তরটা যে অনেক সহজ হত সেটা বুঝতেই পারছ--

$$\begin{aligned}
\widehat{T} \in (-c, c) &\iff X - np \in \big(-c\sqrt{n\widehat{p}(1-\widehat{p})}, c\sqrt{n\widehat{p}(1-\widehat{p})}\big) \\
&\iff np \in \big(X - c\sqrt{n\widehat{p}(1-\widehat{p})}, X + c\sqrt{n\widehat{p}(1-\widehat{p})}\big) \\
&\iff p \in \left(\widehat{p} - c\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \widehat{p} + c\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right).
\end{aligned}$$

তার মানে confidence interval-টা দাঁড়ালো

$$\left(\widehat{p} - c\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \widehat{p} + c\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right)..$$

আমাদের আগের উত্তরের তুলনায় এই উত্তরটা কত সহজ আর মিষ্টি, তাই না?

Example 18: A poll is taken in which 592 out of 985 randomly selected voters indicate their preference for a certain candidate. Construct a 95% confidence interval to estimate the true proportion of voters who prefer the candidate.[Given that $\frac{1}{\sqrt{2\pi}}\int_{1.96}^{\infty} e^{-x^2/2}dx = 0.025$.][5] **(2013.5c)**
SOLUTION: এখানে দুটো গল্প আছে, প্রথমটা বলেছে, দ্বিতীয়টা আমাদের কল্পনা করে নিতে হবে। গল্প দুটো বুঝে নিই--

- প্রথম গল্পটা হল--একটা জায়গায় ভোট হবে, আমরা আগেভাগে বোঝার চেষ্টা করছি একজন বিশেষ প্রার্থী কত ভোট পাবে। তার জন্য আমরা 985 জন ভোটারকে পাকরাও করে জিজ্ঞেস করেছি--"দাদা (বা দিদি), আপনি কি আমাদের সেই বিশেষ প্রার্থীকে ভোট দেবেন গো?" এই 985 জনকে random-ভাবে নেওয়া হয়েছে (যেমন, সেই প্রার্থীর শ্বশুরবাড়ির পাড়ার লোকদেরই খালি জিজ্ঞেস করেছি, এমন নয়)। এদের মধ্যে 592 জন বলেছে যে তারা ওই বিশেষ প্রার্থীকেই ভোট দেবে।
- দ্বিতীয় গল্পটা এইরকম--বিপুলসংখ্যক ভোটার আছে, যাদের মধ্যে কেউ কেউ সেই বিশেষ প্রার্থীকে ভোট দেবে, বাকীরা দেবে না। আমরা ধরে নিচ্ছি যে, এদের মধ্যে যে কোনো ভোটারকে যদি জিজ্ঞাসা করা হয় যে, তিনি বিশেষ প্রার্থীকে ভোট দেবেন কিনা, তবে তিনি সত্য কথাটা জানিয়ে দিতে দ্বিধা করবেন না। এইটা ধরে নেওয়াটা অবশ্যই অগণতান্ত্রিক, নইলে আর গোপনে ব্যালট ব্যবহার করে ভোট হয় কেন? আমাদের 985 জনকে ধরে নেব সেই বিপুলসংখ্যক ভোটার জনতার প্রতিনিধি।

যাই হোক, এই দুই গল্পের statistical নির্যাস হল এই--একজন random ভোটারকে প্রশ্ন করে উত্তর পাওয়া হল আমাদের random experiment. এর দুটো outcome– হ্যাঁ বা না, ঠিক যেমন কয়েন টসের বেলায় head আর tail. যেহেতু প্রচুর সংখ্যক ভোটার আছে, তাই আমাদের 985 জনকে আমরা সেই একই random experiment-এর 985-খানা IID trial ধরে নেব।

> We consider each response as a coin toss, where a "head" denotes willingness to vote for the particular candidate.
> Let $X =$ the total number of heads.
> We assume that the responses are IID (assuming that population of all voters is very large).
> Here the total number of trials is $n = 985$ and $X = 592$.
> Let p be the true proportion who prefer the particular candidate.

মানে, যদি সেই বিপুলসংখ্যক যাবতীয় ভোটারদের প্রত্যেককে প্রশ্নটা করা হত, তবে কতজন "হ্যাঁ" বলত। যেহেতু বিপুলসংখ্যক ভোটারদের প্রত্যেককে তো আর প্রশ্ন করে দেখা সম্ভব হয় নি, তাই p আমাদের অজানা। এই p-এর confidence interval বার করাই আমাদের উদ্দেশ্য।
আমরা প্রথমে অতি সংক্ষেপে লিখে নেব confidence interval-এর ফর্মুলাটা কী এবং কোথা থেকে এল।

> So $X \sim \text{Binomial}(n, p)$.
> We know from CLT that
> $$T = \frac{X - np}{\sqrt{np(1-p)}}$$
> is asymptotically $N(0, 1)$.
> Also, by WLLN, $\widehat{p} = \frac{X}{n} \xrightarrow{P} p$. So
> $$T' = \frac{X - np}{\sqrt{n\widehat{p}(1-\widehat{p})}}$$
> is also asymptotically $N(0, 1)$.

এখানেই আমরা ছোট্টো করে Slutsky's theorem লাগালাম, যদিও খটমট নামটা উল্লেখ করে জটিলতা বৃদ্ধি করি নি।

So $P(-1.96 < T' < 1.96) = 1 - 2 \times 0.025 = 0.95$. Hence a possible 95% level confidence interval for p is

$$\left(\widehat{p} - 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}, \widehat{p} + 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right).$$

In our case, $\widehat{p} = \frac{X}{n} = \frac{592}{985} = 0.601$. So the confidence interval is $(0.57, 0.63)$.

এই প্রসঙ্গে সেই বদখত confidence interval-টা এক্ষেত্রে কী হয় সেটাও বার করে দেখতে পারো। তার জন্য আমাদের এই quadratic equation-টার root দুটো বার করতে হচ্ছিল--

$$(n^2 + c^2)p^2 - (2nX + c^2)p + X^2 = 0.$$

এখানে $n = 985, X = 592$ আর $c = 1.96$. তাই root দুটো আসছে 0.600 আর 0.602. সুতরাং বদখত ফর্মুলাটা লাগালে যে confidence interval-টা পাচ্ছি সেটা হল $(0.600, 0.602)$, যেটা নিঃসন্দেহে অনেক বেশী সুক্ষ্ম। ■

Exercise 4: 171 out of 300 voters picked at random from a large electorate said that they are going to vote a particular candidate. Find 95% confidence interval for the proportion of voters of the electorate who would vote in favour of the candidate. (Given $P(U \geq 1.96) = 0.025$.) [7] **(2004.10b)**

Hint:

এটাও আগের অংকটার মতই। তোমার হাত পাকানোর জন্য রেখে দিলাম। এখানে শেষে যে $P(U \geq 1.96) = 0.025$ লিখেছে ওটা আসলে অসম্পূর্ণ। বলতে চেয়েছে যে যদি $U \sim N(0,1)$ হয়, তবে $P(U \geq 1.96) = 0.025$ হবে। মানে $\Phi^{-1}(1 - 0.025) = 1.96$, যেখানে $\Phi(x)$হল $N(0,1)$-এর CDF. ■

Answers

4. $(0.514, 0.626)$.

Chapter X
Test of Hypotheses

DAY 35
গোড়ার কথা

35.1 নামের নামতা

একজনের মাথার ব্যারাম ছিল, সে সব জিনিসের নামকরণ করত। তার জুতোর নাম ছিল অবিমৃষ্যকারিতা, তার ছাতার নাম ছিল প্রত্যুৎপন্নমতিত্ব, তার গাড়ুর নাম ছিল পরমকল্যাণবরেষু--কিন্তু যেই তার বাড়ির নাম দিয়েছে কিংকর্তব্যবিমূঢ় অমনি ভূমিকম্প হয়ে বাড়িটাড়ি সব পড়ে গিয়েছে।

--হ য ব র ল (সুকুমার রায়)

বিজ্ঞানের বিভিন্ন শাখায় আমাদের অনেক সময়ে সিদ্ধান্ত করতে হয় যে, দুটো জিনিসের মধ্যে কোনটা ঘটেছে, কিন্তু সেটা সরাসরি জেনে ফেলার কোনো উপায় থাকে না। তখন অন্যান্য তথ্যের ভিত্তিতে আন্দাজে সিদ্ধান্ত করতে হয়। যেমন ধরো, আমার হাতে একটা টিউমার হয়েছে, আমি ভয়ে ভয়ে ডাক্তারের কাছে গিয়ে জিজ্ঞাসা করছি যে, ক্যান্সার-ট্যান্সার হয়েছে কিনা। সুতরাং এখানে দুটো জিনিস হল--এক, ক্যান্সার হয়েছে, আর দুই, ক্যান্সার হয় নি। এদের প্রত্যেকটাকে বলে একেকটা **hypothesis.** ডাক্তারবাবু এবার কি করবেন? তিনি কিছু পরীক্ষা করে আনতে বলবেন। এই পরীক্ষার রিপোর্ট হল আমাদের data. তুমি আমি সেই রিপোর্টের মাথামুণ্ডু কিছুই বুঝব না, আমাদের অনভিজ্ঞ চোখে খালি একরাশ সংখ্যার স্তূপ বলে মনে হবে। কিন্তু ডাক্তারবাবু ওই সংখ্যার স্তূপ থেকে ঠিক দরকারী সংখ্যাটুকু ছেঁকে তুলে নেবেন। এই সার নির্যাসটাকে বলে **test statistic.** ধরো এই দরকারী সংখ্যাটা আমার ক্ষেত্রে হয়েছে 2.34. এই সংখ্যাটা বলে দিলেও তুমি আমি বুঝব না সেটা ভালো নাকি খারাপ খবর। ডাক্তারবাবু অভিজ্ঞ মানুষ, তিনি জানেন যে, এই সংখ্যাটা যত বেশী হবে তত ক্যান্সারের সম্ভাবনা। কতটা বেশী হলে সেটা বেশ বিপজ্জনক বলে ধরা যায়? ডাক্তারীশাস্ত্রে তারও বিধান আছে। ধরো 2.5-এর নীচে থাকলে ব্যাপারটা তত ভয়ের কিছু নয়। এখন, আমার ক্ষেত্রে যেহেতু $2.34 < 2.5$, সুতরাং ডাক্তারবাবু আমাকে অভয় দিয়ে বলবেন-- আপনি নাচতে নাচতে বাড়ি চলে যান, চিন্তার কিচ্ছু নেই! এই যে বিপদসীমা 2.5 সংখ্যাটা, এটাকে বলে **critical value.**
এখন কথা হল, ডাক্তারীশাস্ত্রের বিধান তো আর সব ক্ষেত্রে অভ্রান্ত হয় না। সুতরাং ডাক্তারবাবুর যতই হাতযশ থাক, ভুল হতেই পারে। ভুলটা হতে পারে দুই রকম--

- এক, আমার ক্যান্সার হয়েছে, কিন্তু ডাক্তারবাবু সেটা ধরতে পারেন নি।
- আর দুই, আসলে ক্যান্সার হয় নি, অথচ ডাক্তারবাবু রায় দিয়েছেন ক্যান্সার হয়েছে বলে।

আশা করি সন্দেহ নেই যে প্রথম ভুলটা অনেক বেশী মারাত্মক। একে আমরা বলব **type I error.** অন্য ভুলটাকে বলব **type II error.** মনে রেখো যে, বেশী মারাত্মক ভুলটাকে বলে type I error. ভুল মানে হল ডাক্তারের রায়ের সঙ্গে সত্য মেলে নি। Type I error-এর বেলায় সত্য যেটা তাকে বলব **null hypothesis.** আমাদের উদাহরণে type I

error মানে ক্যান্সার হয়েছে, অথচ ডাক্তারবাবু বলেছেন হয় নি। সুতরাং সত্য হল "ক্যান্সার হয়েছে"। এইটাই আমাদের null hypothesis. অন্য সম্ভাবনাটাকে বলব **alternative hypothesis.** যারা অন্য বই থেকে আগেই খানিকটা পড়েছো, তারা এইখানে একটু সাবধান। অধিকাংশ বইতেই আগে null আর alternative hypothesis-এর সংজ্ঞা দিয়ে, তারপর তাদের ভিত্তিতে দুই ধরণের error-এর সংজ্ঞা দেয়। কিন্তু বহু পাঠ্যপুস্তকে সেভাবে করলেও সেটা সঠিক সংজ্ঞা নয়, কারণ বাস্তবে hypothesis দুটোর কপালে সাইনবোর্ড আঁটা থাকে না null আর alternative বলে। যেটা জানা থাকে, সেটা হল কোন error-টা বেশী মারাত্মক, তার ভিত্তিতে ঠিক করতে হয় কোন hypothesis-টাকে null বলব, আর কোনটাকে alternative. সাধারণতঃ null hypothesis-কে লোকে H_0 আর alternative hypothesis-কে H_1 লেখে। তার মানে আমাদের উদাহরণে

H_0 : "ক্যান্সার হয়েছে" বনাম H_1 : "ক্যান্সার হয় নি।"

যদি ডাক্তারবাবু বলেন "ক্যান্সার হয়েছে" তবে সেটাকে বলব **"accepting H_0".** যদি ডাক্তারবাবু বলেন "ক্যান্সার হয় নি" তবে সেটা হবে **"rejecting H_0".**

ডাক্তারবাবু যদি ওস্তাদ হন, তবে আশা করা যায় যে দুরকম ভুলের সম্ভাবনাই কম থাকবে, বিশেষতঃ type I error-এর probability-টা কম থাকাটা খুবই জরুরী, কারণ ওইটাই বেশী বিপজ্জনক। সেই কারণে ডাক্তারীশাস্ত্রে যে সব পরীক্ষার বিধান আছে সেগুলো এমনভাবে তৈরী করা হয়, যাতে type I error-এর probability কোনোভাবেই একটা নির্দিষ্ট সীমার উপরে না যায়। এই ঊর্দ্ধ সীমাকে বলা হয় **level of significance.** যেমন, যদি শোনো যে level of significance হল 0.05, তার মানে যদি 100-টা রোগী আসে যাদের টিউমারে সত্যিই ক্যান্সার আছে, তবে গড়ে বড় জোর 5-টা কেসে ডাক্তারবাবু ভুল করবেন।

Example 1: শুধু ডাক্তারীশাস্ত্রেই নয়, আমরাও প্রতিদিনকার জীবনে test of hypothesis-এর ধারণাটা ব্যবহার করে থাকি। যেমন ধরো তোমার হাতে একটা বন্ধ কৌটো আছে। তুমি বুঝতে চাইছ ওর মধ্যে জল আছে কিনা। সেটা বোঝার একটা কায়দা হল কৌটোটাকে ঝাঁকিয়ে দেখা কীরকম শব্দ হয়। এই সহজ ব্যাপারটার মধ্যে কীভাবে test of hypothesis লুকিয়ে আছে দ্যাখো--

সহজ ভাষায়	test of hypothesis-এর ভাষায়
কৌটোয় হয় জল আছে, ...	H_0
... নয়তো জল নেই	H_1
কৌটো ঝাঁকালাম	data collection
কীরকম শব্দ হয় শুনলাম	test statistic
জল হলে কীরকম শব্দ হত?	distribution of test statistic under H_0. এর আরেক নাম **null distribution.**
আমাদের শব্দটা কি সেইরকম?	critical value ব্যবহার করে **null distribution**-এর সাথে তুলনা
যদি না হয়, তবে কৌটোয় জল নেই।	reject H_0
যদি হয়, তবে জল হতে পারে	accept H_0

এই উদাহরণে দুটো নতুন ধারণার কথা এসেছে--

- Null distribution হল H_0 ঠিক হলে test statistic-এর আচরণ কীরকম হত, তার হিসেব।
- যদি H_0-টা reject করি, তবে সেটা জোর দিয়ে করা যায়, কারণ জলের আওয়াজ জলের মতই হওয়া উচিত। কিন্তু যদি H_0-কে accept করা হয়, তবে সে জোরটা থাকে না, কারণ কোনো কিছুর আওয়াজ জলের মত হলেই সেটা জল হবে এমন তো কথা নেই, দুধও হতে পারে। সেই যেমন রবীন্দ্রনাথেরও দাড়ি আছে, ছাগলেরও দাড়ি আছে, কিন্তু তা বলে...

■

এইবার একই উদাহরণের সামান্য রূপভেদ দেখব, যাতে বোঝা যাবে H_0-এর উপরে কীভাবে test statistic-টা নির্ভর করে।

Example 2: আবার সেই কৌটোর সমস্যা। কিন্তু এবার বুঝতে হবে কৌটোয় তুলো আছে কিনা। এবার অবশ্যই আওয়াজ শোনার চেষ্টা করে লাভ নেই। এবার দেখব ওজন।

কৌটোয় হয় তুলো আছে, ...	H_0
... নয়তো তুলো নেই	H_1
কৌটো ঝাঁকালাম	data collection
কীরকম ওজন অনুভব করলাম	test statistic
তুলো হলে কীরকম ওজন হত?	null distribution of test statistic
কৌটোর ওজনটা কি সেইরকম?	critical value ব্যবহার করে null distribution-এর সাথে তুলনা
যদি না হয়, তবে কৌটোয় তুলো নেই।	reject H_0
যদি হয়, তবে তুলো হতে পারে	accept H_0

■

35.2 অংকের ভাষায়

এইবার আলোচনাটাকে অংকের ভাষায় লিখব। তার জন্য একটা অন্য উদাহরণ মাথায় রাখলে সুবিধা হবে। এই বইতে ক্যাসিনোর উল্লেখ বারবারই এসেছে। সেখানকার প্রধান আকর্ষণ হল বিভিন্ন random experiment-এর উপর বাজি ধরা। বুঝতেই পারছ যে, ক্যাসিনোর মালিকের চেষ্টা থাকে random experiment-গুলোর মধ্যে জাল জুয়াচুরি ঢোকানোর, যেমন যে কয়েনটাকে unbiased বলে দাবী করা হচ্ছে সেটা হয়তো আসলে biased, এইরকম। এইসব নিয়ে অনেক টাকার মামলা মোকদ্দমা চলে, এবং বড় বড় ক্যাসিনোর মালিকরা অনেক সময়ে আগেভাগেই নিরপেক্ষ বিশেষজ্ঞ ডেকে এনে random experiment-এর যন্ত্রপাতি পরীক্ষা করে সার্টিফিকেট করিয়ে রাখে। এরই একটা সহজ সংস্করণ নিয়ে এই উদাহরণটা।

Example 3: একটা কয়েন আছে। দাবী করা হয়েছে যে, সেটা unbiased. কিন্তু সে বিষয়ে সন্দেহের অবকাশ থাকায় বিশেষজ্ঞ দল এসেছে। এখানে $P(\text{head})$-কে যদি θ বলি, সেটা হল একটা unknown parameter. আমরা জানি না θ কত, কিন্তু এটুকু জানি যে θ হল $(0,1)$-এর মধ্যে একটা সংখ্যা, অর্থাৎ কিনা $(0,1)$ হল parameter space. এর নাম দিলাম ধরো Θ. লক্ষ কর যে, এখানে θ কত সেটা গুরুত্বপূর্ণ নয়। গুরুত্বপূর্ণ হল খালি এটুকু আন্দাজ করতে পারা যে, $\theta = \frac{1}{2}$ কিনা। এর ফলে parameter space-টা দুভাগে ভাগ হয়ে গেল, $\{\frac{1}{2}\}$ আর $\left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right)$. Θ-র এই দুটো subset-কে আমরা যথাক্রমে Θ_0 এবং Θ_1 নাম দেব। লক্ষ কর যে--

$$\Theta = \Theta_0 \cup \Theta_1 \text{ এবং } \Theta_0 \cap \Theta_1 = \phi.$$

এর মধ্যে প্রথমটা বলছে যে, কয়েনটা হয় unbiased নয়তো biased, এ দুয়ের বাইরে কিছু হতে পারে না। দ্বিতীয়টা বলছে যে, একটা কয়েন একই সঙ্গে unbiased এবং biased হতে পারে না। এখানে আমাদের hypothesis দুটো হল

$$H_0: \text{ unbiased এবং } H_1: \text{ biased.}$$

এদেরকে আমরা অংকের ভাষায় লিখতে পারি--

$$H_0 : \theta \in \Theta_0 \text{ এবং } H_1 : \theta \in \Theta_1.$$

এই দুটো hypothesis-এর মধ্যে কোনটা সত্য, সেটা নির্ণয় করার জন্য ধরো বিশেষজ্ঞরা ঠিক করলেন যে, কয়েনটাকে 100 বার টস করে দেখা হবে। আমরা এই 100-টা টসের ফলাফলগুলোকে $X_1, X_2, ..., X_{100}$ লিখব। যেমন X_1 হয়তো "head", X_2 হয়তো "tail", এইরকম। এইটা হল আমাদের data. তারপর আমরা গুণে দেখব মোট কতগুলো head পড়েছে। ধরো

এই সংখ্যাটার নাম দিলাম T. চট্ করে বুঝে নাও যে, T হল $X_1, ..., X_{100}$-এর একটা function. অর্থাৎ কিনা T হল একটা statistic. যেহেতু test-এর কাজে ব্যবহার করা হবে, তাই এই T-কে আমরা বলব test statistic. এবার মনে করো বিশেষজ্ঞরা বললেন--

> যদি T-টা 40 থেকে 60-এর মধ্যে থাকে, তবে কয়েনটাকে unbiased বলে মেনে নেওয়া হবে, নইলে bias আছে বলে সন্দেহ করব।

অর্থাৎ--

> $T \in \{40, 41, ..., 60\}$ হলে আমরা H_0-এর স্বপক্ষে রায় দেব (মানে H_0-কে accept করব)। নইলে H_0-কে reject করব।

এখানে দুটো critical value রয়েছে 40 আর 60. আর $X_1, ..., X_{100}$-এর যে সব value-র জন্য আমরা H_0-কে reject করব (মানে যখন $T < 40$ বা $T > 60$ হবে) তাদের set-কে বলব **critical region.** ■

লক্ষ করো যে এখানে আমরা তিন ধরণের জিনিস নিয়ে কাজ করছি--

- θ, যেটা fixed কিন্তু অজানা,
- Θ, Θ_0 এবং Θ_1 যারা তিনটে fixed set এবং জানা
- $X_1, ..., X_{100}$ এবং T, এরা random (কারণ কোন টসে কী এসেছে তার উপর নির্ভর করে), কিন্তু এদের value-গুলো আমাদের জানা।

যেহেতু আমাদের সিদ্ধান্ত (বা বলা ভালো অনুমান) নির্ভর করছে T-এর উপরে, এবং T নিজেই হল random, তাই আমাদের সিদ্ধান্তও random. সুতরাং type I বা type II error হচ্ছে কিনা সেটাও random. তাদের probability আমরা লিখতে পারি এইভাবে--type I error-এর probability হল

$$P(H_0 \text{ is rejected}) \text{ যখন } H_0 \text{ ঠিক},$$

আর type II error-এর

$$P(H_0 \text{ is accepted}) \text{ যখন } H_1 \text{ ঠিক},$$

আমরা বলব যে, এই test-টার level of significance হল α যদি type I error-এর probability সর্বদাই $\leq \alpha$ হয়, মানে

$$\forall \theta \in \Theta_0 \quad P(H_0 \text{ is rejected}) \leq \alpha.$$

এই প্রসঙ্গে আরেকটা নামও শিখে রাখো--একটা test-এর **power** মানে হল $1 - P(\text{type II error})$.
একরাশ নাম শেখা হল। একটা ছোটো উদাহরণ করে রাখলে সবকিছু মনে রাখার সুবিধা হবে। এই উদাহরণটা উপরের উদাহরণটার মতই, কেবলমাত্র নামগুলো মনে রাখানোর জন্য তৈরী।

Exercise 1: আবার একটা কয়েন নিলাম। এবার আমরা জানতে চাই সেটাতে head পড়ার probability-টা $\leq \frac{1}{2}$ নাকি $> \frac{1}{2}$. এখানে H_0 নিচ্ছি "probability $\leq \frac{1}{2}$" আর H_1 হল "probability $> \frac{1}{2}$"। সিদ্ধান্ত করার জন্য আমরা কয়েনটাকে দুইবার টস করলাম। যদি দুবারই head পড়ে, তবে রায় দেব $> \frac{1}{2}$, নইলে $\leq \frac{1}{2}$. এই বর্ণনা থেকে সনাক্ত করো $\theta, \Theta, \Theta_0, \Theta_1$ কী কী। এখানে data-ই বা কী, আর test statistic-ই বা কোনটা? Critical region-টাও বার কোরো। ■

এতক্ষণ যা যা শিখলাম, এবার সে সব গুছিয়ে লেখার সময় এসেছে।

Example 4: What is meant by a statistical hypothesis?[1] **(2014,2011,2010,2009,2007,2004)**

SOLUTION:

> Suppose that we have a data set $X_1, ..., X_n$ with joint distribution determined by an unknown parameter θ, taking value in a known parameter space Θ. Then any statement of the form $\theta \in \Theta'$ is called a statistical hypothesis where Θ' is some known subset of Θ.

একটা উদাহরণ দিয়ে নেওয়া ভালো--

> For example, if a coin with unknown $P(\text{head})$ is tossed 10 times, then the $P(\text{head})$ is an unknown parameter taking value in $(0, 1)$. The statement $P(\text{head}) = \frac{1}{2}$ is a statistical hypothesis.

■

এই উত্তরের ভাষাটা খেয়াল রেখো। একই জিনিস বারবার করে নীচের অংকগুলোতেও লাগবে।

Example 5: What is hypothesis testing?[2] **(2013.6b)**

SOLUTION: গোড়ার গল্পটা আগেরই মত--

> Suppose that we have a data set $X_1, ..., X_n$ with joint distribution determined by an unknown parameter θ, taking value in a known parameter space Θ. Then we often want to decide which of the following two hypotheses is true:
>
> $$H_0 : \theta \in \Theta_0 \text{ and } H_1 : \theta \in \Theta_1,$$
>
> where Θ_0, Θ_1 are two known subsets of Θ with $\Theta_0 \cap \Theta_1 = \phi$ and $\Theta_0 \cup \Theta_1 = \Theta$.

এখানে চট্ করে বলে রাখি যে, "hypotheses" হল "hypothesis"-এর বহুবচন।

> Hypothesis testing means constructing some rule based on data to make this decision (*i.e.*, either accept H_0 or reject it).

■

এবার আমরা আরও দুটো নতুন ভাষা শিখব--কখন একটা hypothesis-কে simple বলে আর কখন তাকে composite বলে।

Example 6: Explain the following terms: Simple hypothesis and composite hypothesis Illustrate the answer with examples.[2] **(2004)**

SOLUTION: গোড়ার গল্পটা তো সেই আগেরই মত--

> Suppose that we have a data set $X_1, ..., X_n$ with joint distribution completely determined by an unknown parameter θ, taking value in a known parameter space Θ. Then any statement of the form $\theta \in \Theta'$ is called a statistical hypothesis where Θ' is any known subset of Θ.

> If Θ' is a singleton set (*i.e.*, the joint distribution is completely known if $\theta \in \Theta'$, then the hypothesis is called a simple hypothesis, otherwise it is called a composite hypothesis.

একটা উদাহরণ হয়ে যাক--

> For example, if $X_1, ..., X_n$ are iid $N(\mu, \sigma)$ where $\mu \in \mathbb{R}$ and $\sigma > 0$ are both unknown then the hypothesis $\mu = 0$ is a composite hypothesis, because it is actually $\theta \equiv (\mu, \sigma) \in \{0\} \times (0, \infty) \equiv \Theta_0$. However, if σ is known, then $\mu = 0$ is a simple hypothesis.

এই উদাহরণটা ভালো করে বুঝে নাও। ■

Example 7: Explain the following terms: Null hypothesis and alternative hypothesis. Illustrate the answer with examples.[2] **(2004)**

SOLUTION:

> In statistical hypothesis testing, we deal with a pair of hypotheses of the form
>
> $$\theta \in \Theta_0 \text{ and } \theta \in \Theta_1,$$
>
> where Θ_0, Θ_1 are two known subsets of Θ with $\Theta_0 \cap \Theta_1 = \phi$ and $\Theta_0 \cup \Theta_1 = \Theta$. One of these is called the null hypothesis depending on the following points:
>
> 1. wrong rejection of the null hypothesis is more dangerous than wrong rejection of the other,
> 2. it is the hypothesis that we would accept in absence of convincing evidence against it.

সহজ ভাষায় বলতে গেলে এই--যদি দ্যাখো কোনো পক্ষেই খুব জোরালো যুক্তি নেই, তবে তুমি সেই hypothesis-এর স্বপক্ষেই রায় দেবে যেটা বেশী নিরাপদ। সেই নিরাপদ hypothesis-টাকেই বলে null hypothesis. আমরা আগেই বলেছিলাম যে type I error হল বেশী মারাত্মক। যে hypothesis-এর স্বপক্ষে রায় দিলে ভুল হলেও অন্ততঃ type I error হবে না সেটাই H_0.

> The other hypothesis is called the alternative hypothesis.

এবার কিছু উদাহরণ চাই--

> For example, if we are testing a tumour for cancer, the two hypotheses are ``cancer'' and ``not cancer''. Here the first one will be taken as the null hypothesis, because failure to diagnose cancer is more dangerous than

> falsely diagnosing cancer.

■

Exercise 2: Define (i) a simple hypothesis, (ii) a composite hypothesis, (iii) a null hypothesis.[3] **(2010.5a)**
HINT: আগেই দুটো অংকেই করা আছে। ■

Example 8: What are the types of error of decision committed in testing of statistical hypothesis?[2] **(2011,2009,2007)**
SOLUTION:

> In statistical testing of some null hypothesis, H_0, against some alternative hypothesis, H_1, there are two possible errors:
>
> 1. Type I error: Rejecting H_0 when it is actually true.
> 2. Type II error: Accepting H_0 when it is actually false.

উদাহরণ দিতে ভুলো না--

> For example, if we are testing a tumour for cancer, and we have H_0 : "cancer" and H_1 : "not cancer". Then type I error will be committed if the tumour is actually cancerous, but the doctor fails to diagnose it. Type II error will be committed, if the doctor wrongly diagnoses cancer, when the tumour is actually harmless.
>
> Type I error is generally more harmful than type II error.

■

Example 9: Explain type I and type II errors with suitable examples.[3] **(2013.6b)**
SOLUTION: আগের অংকেই বলেছি। ■

Example 10: Explain the concept of critical region and power of a test.[2] **(2011,2009,2007,2004)**
SOLUTION:

> In a statistical test of hypothesis we have a data set that can take values in some known sample space $\mathscr{X}$.

এই "sample space" ব্যাপারটা বুঝে নাও। আমাদের data-র সম্ভাব্য সবরকম value-র set-টাকে বলে **sample space.** যেমন যদি দুবার কয়েন টস করে তার outcome দুটো নিয়ে আমাদের data তৈরী হয়, তবে sample space হবে

$$\{\ HH,\ HT,\ TH,\ TT\ \}.$$

> Also we have two hypotheses--the null hypothesis, H_0, and the alternative hypothesis, H_1. Exactly one of the two hypotheses is true, but we do not know which. A statistical test is a rule that tells us for each possible value of the data set whether to accept H_0 or to reject it. The subset of $\mathscr{X}$ for which the test rejects H_0 is called the critical region of the test.

একটা উদাহরণ--

> For example, suppose that
>
> - our data set consists of the outcomes of three tosses of a coin with $P(\text{head}) = \theta \in (0,1)$ unknown.
> - $H_0 : P(\text{head}) \geq \frac{1}{2}$ and $H_1 : P(\text{head}) < \frac{1}{2}$,
> - we reject H_0 if and only if we get ≤ 1 heads.
>
> Then the sample space is
>
> $$\mathscr{X} = \{\ HHH,\ HHT,\ HTH,\ HTT,\ THH,\ THT,\ TTH,\ TTT\ \}$$
>
> and the critical region is
>
> $$\{\ HTT,\ THT,\ TTH,\ TTT\ \}$$

এরা হচ্ছে সেইসব কেস এখানে head-এর সংখ্যা ≤ 1.

> The probability of rejecting H_0 when it is actually false is called the power of a test. It is also $1 - P(\text{type II error})$. Generally, power is expressed as a function of the unknown parameter(s).

এটারও একটা উদাহরণ দিয়ে দিই--

> In the above example,
>
> $$\text{power}(\theta) = 3\theta(1-\theta)^2 + (1-\theta)^3, \quad \theta \in \left(0, \tfrac{1}{2}\right).$$

এইটা হচ্ছে critical region-টার probability. ■

Example 11: What is the power of a test?[2] **(2013.6b)**

SOLUTION: আগের অংকেই দেওয়া আছে। ■

35.3 হাতে কলমে

Example 12: Let p denote the probability of getting a head when a given coin is tossed once.

Suppose that the hypothesis $H_0 : p = 0.5$ is rejected in favour of $H_1 : p = 0.6$ when the result is 7 or more heads in tossing a coin 10 times. Calculate the probability of type I and type II error.[6] **(2005.10b)**

SOLUTION: ব্যাপারটা বুঝে নেওয়া যাক। একটা কয়েন আছে। কোনো যাদুমন্ত্রবলে আমরা জেনে গিয়েছি যে তার $P(\text{head})$ হল হয় 0.5 নয়তো 0.6. মানে $p \equiv P(\text{head})$ হল আমাদের parameter, আর parameter space হল $\{0.5, 0.6\}$. এখন আমরা ঘোরতর দুশ্চিন্তায় পড়ে গিয়েছি যে--আসলে $P(\text{head}) = 0.5$ নাকি $P(\text{head}) = 0.6$? সেই গুরুতর সমস্যার নিষ্পত্তি করতে আমরা কয়েনটাকে টস করেছি 10 বার। এমন সময়ে আকাশ থেকে মেঘমন্দ্র স্বরে দৈববাণী হয়েছে--হে বৎস, যদি দ্যাখো ≥ 7 খানা head পড়েছে, তবে তুমি $p = 0.6$-এর স্বপক্ষে রায় দিও। শুনে আমরা হকচকিয়ে গিয়ে ভাবছি, দৈববাণী মানতে গিয়ে শেষে ভুল করব না তো? সেই জন্যই দুইধরণের error-এর probability বার করে রাখতে চাইছি। প্রথমে set up-টা লিখে নিই--

> Data: X = number of heads out of 10 IID tosses.
>
> Then we know $X \sim \text{Binomial}(10, p)$, where $p = P(\text{head})$ is our parameter.
>
> Parameter space: $\{0.5, 0.6\}$.
>
> To test:
>
> $$H_0 : p = 0.5 \text{ against } H_1 : p = 0.6.$$
>
> Test procedure:
>
> $$\phi(X) = \begin{cases} 1 & \text{if } X \geq 7 \\ 0 & \text{otherwise} \end{cases}$$

এখানে একটা নতুন কায়দা ব্যবহার করেছি লেখার জন্য। একটা test-কে আমরা একটা function হিসেবে লিখতে পারি, $\phi(X)$, যেখানে $\phi(X) = 1$ মানে হল "reject H_0" আর $\phi(x) = 0$ মানে হল "accept H_0."

> Then
>
> $$\begin{aligned} P(\text{type I error}) &= P(\text{rejecting } H_0) \text{ when } H_0 \text{ is true} \\ &= P(X \geq 7) \text{ when } p = 0.5 \\ &= \left[\binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right] 0.5^{10} \\ &= 0.172. \end{aligned}$$

Also

$$\begin{aligned} P(\text{type II error}) &= P(\text{accepting } H_0) \text{ when } H_1 \text{ is true} \\ &= P(X < 7) \text{ when } p = 0.6 \\ &= 1 - P(X \geq 7) \text{ when } p = 0.6 \\ &= 1 - \left[\binom{10}{7} 0.6^7 0.4^3 + \binom{10}{8} 0.6^8 0.4^2 + \binom{10}{9} 0.6^9 0.4^1 + \binom{10}{10} 0.6^{10} 0.4^0\right] \\ &= 0.618. \end{aligned}$$

■

Exercise 3: Let p denote the probability of getting a head when a given coin is tossed once. Suppose that the hypothesis $H_0 : p = \frac{1}{2}$ is rejected in favour of $H_1 : p = \frac{3}{4}$ if more than 3 heads are obtained out of 5 throws of a coin. Find the probabilities of type I and type II errors.[6] **(2007.5b)**
HINT: আগের অংকের অনুকরণে করো দেখি।
■

এই যে দুটো অংক করলাম, এখানে H_0এবং H_1 দুজনেই ছিল simple. এবার একটা অংক দেখব যেখানে H_1 হবে composite. এইরকম ক্ষেত্রে P(type II error) আর power-কে parameter-এর function হিসেবে লিখতে হয়।

Example 13: Given the population density function $f(x, \theta) = \theta e^{-\theta x}$, $0 \leq x < \infty$, $\theta > 0$. The null hypothesis $H_0 : \theta = 2$ against the one sided alternative $H_1 : \theta > 2$ will be tested on the following test procedure: H_0 should be rejected if a sample drawn from the population (x) is greater than or equal to 6. Find the type I error and power of the test.[6] **(2006.5c)**
SOLUTION: এই অংকের ভাষায় কিছু ত্রুটি আছে--

- যেহেতু $\theta > 0$ লিখেছে, তাই মনে হচ্ছে যেন parameter space এখানে $\Theta = (0, \infty)$. ওদিকে H_0 আর H_1-এর বর্ণনা থেকে দেখছি $\Theta_0 = \{2\}$ আর $\Theta_1 = (2, \infty)$. সুতরাং এখানে $\Theta = \Theta_0 cup \Theta_1$ হচ্ছে না। আসলে এখানে $\Theta = [2, \infty)$ নিতে হবে, মানে $\theta > 0$ না বলে $\theta \geq 2$বলা উচিত ছিল।
- প্রশ্নটায় "population" শব্দটার পর (x) লিখেছে। আসলে (x) লেখা উচিত ছিল "sample" শব্দটার পরে।

যাই হোক, আমরা অংকটাকে মেরামত করে নিয়ে সমাধান করব। Set up-এর বর্ণনা দিয়ে শুরু করি--

Data: X

Given that X has density

$$f(x, \theta) = \begin{cases} \theta e^{-\theta x} & \text{if } 0 \leq x < \infty \\ 0 & \text{otherwise} \end{cases}$$

Here θ is our unknown parameter.

Parameter space: $[2, \infty)$.

এটা পেয়েছি $H_0 : \theta = 2$ আর $H_1 : \theta > 2$-এর union নিয়ে।

To test:

$$H_0 : \theta = 2 \text{ against } H_1 : \theta > 2$$

এবার test-টা লিখব একটা function হিসেবে--

Test procedure:

$$\phi(X) = \begin{cases} 1 & \text{if } X \geq 6 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\begin{aligned} P(\text{type I error}) &= P(\text{rejecting } H_0) \text{ when } H_0 \text{ is true} \\ &= P(X \geq 6) \text{ when } \theta = 2 \\ &= \int_6^\infty 2e^{-2x} dx \\ &= e^{-12}. \end{aligned}$$

এবার আমরা power বার করব। সুতরাং এবার ধরে নেব H_1 হল ঠিক। কিন্তু H_1 ঠিক হলেও আমরা θ-র value জানতে পারছি না, যেহেতু H_1 হল composite. খালি এইটুকু জানা যাচ্ছে যে, $\theta > 2$. তাই উত্তরটা $\theta > 2$-এর জন্য θ-র একটা function হিসেবে বেরোবে।

Let us take any $\theta > 2$. Then the power of the test at this θ is

$$\begin{aligned} & P(\text{rejecting } H_0) \text{ when } H_1 \text{ is true} \\ = \ & P(X \geq 6) \text{ when } \theta > 2 \\ = \ & \int_6^\infty \theta e^{-\theta x} dx, \quad \theta > 2 \\ = \ & e^{-6\theta}, \quad \theta > 2. \end{aligned}$$

Thus the required power function is

$$\text{power}(\theta) = e^{-6\theta}, \quad \theta > 2.$$

■

Exercise 4: Given the population density function

$$f(x, \theta) = \theta e^{-\theta x}, \quad 0 \leq x < \infty, \quad \theta > 0,$$

the null hypothesis $H_0 : \theta = 2$ against the one-sided alternative $H_1 : \theta > 2$ will be tested on the following procedure: H_0 should be rejected if a sample drawn from the population (x) is greater than or equal to 6. Find the type I and type II error and power of the test.[5+1+1] **(2012.5c)**

HINT:

আগের অংকটাই। খালি বাড়তি হল $P(\text{type II error})$ বার করা। মনে রেখো যে, $P(\text{type II error}) = 1-\text{power}$. ■

নীচের অংকটা তোমার করার জন্য রেখে দিলাম।

Exercise 5: Given the density function

$$f(x) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \leq x \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

and that you are testing the null hypothesis $H_0 : \theta = 1$ against $H_1 : \theta = 2$ by means of a single observed value X_1, determine the sizes of type I and type II errors if you choose the interval $0.5 \leq x$ as the critical region.[6] **(2014.6c)**

HINT:

এখানে size বলতে probability বুঝিয়েছে।

■

DAY 36
Goodness-of-fit test

আমরা probability আর statistics করার সময়ে প্রায়শঃই unbiased coin বা fair die, এইসব কথা বলে থাকি। স্বভাবতঃই প্রশ্ন ওঠে যে, যেসব coin বা ছক্কা আমরা দেখতে পাই, তারা কী সত্যিই unbiased? ধরো তুমি লুডো খেলছ। ছক্কাতে একটা ছয় পড়ার খুব দরকার, কিন্তু কিছুতেই আর ছয় পড়ছে না। তখন কেমন জানি একটা সন্দেহ হয় ছক্কাটার সব দিক সমান তো, নাকি ছয়ের দিকটা একটু বেশী ভারী, তাই কিছুতেই উপরে আসছে না? বড় বড় ক্যাসিনোতে, যেখানে লক্ষ লক্ষ টাকার লেনদেন হয় জুয়ো খেলায়, সেখানে এই প্রশ্ন আরও গুরুতর আকার নেয়। সেখানে দস্তুরমত অংক কষে পরীক্ষা করা হয় কোনো ফাঁকি আছে কিনা। এদিক ওদিক বেরোলে যে বিরাট আইনকানুনের লড়াই বেঁধে যায়, সে কথা তো গতকালই বলেছি। এই সব রকমের পরীক্ষারই মূলে আছে এক ধরণের test of hypothesis, যাকে বলে goodness-of-fit test. সহজ ভাষায় বলা যায়--যা হওয়া উচিত তার সঙ্গে বাস্তবটা "কত ভালো করে ফিট করছে" তার পরীক্ষা। এই জিনিসটাই এবার আমরা শিখব। কায়দাটা মোটেই কঠিন কিছু নয়। কিন্তু কী করে কায়দাটা এল সেটা অনেকটাই কঠিন। আমরা সেই প্রসঙ্গে যাব না, খালি কায়দাটুকু শিখেই ছেড়ে দেব।

একটা উদাহরণ নিয়ে শুরু করি।

Example 14: A die was thrown 60 times with the following results:

Face	1	2	3	4	5	6
Frequency	6	10	8	13	11	12

Are the data consistent with the hypothesis that the die is unbiased? [Given $\chi^2_{0.01} = 15.09$ for 5 df.] [5] **(2005.10c)**

SOLUTION: কী চাইছে বুঝে নিই। একটা ছক্কা আছে। দাবি করা হয়েছে যে, এটি একটি অতি সচ্চরিত্র সজ্জন ছক্কা, অর্থাৎ ওর প্রতিটা পাশই আসার সম্ভাবনা সমান (মানে $\frac{1}{6}$ করে)। তুমি অবশ্য এইরকম দাবিতে ভোলবার পাত্র নও। তাই তুমি ছক্কাটাকে হাতে নিয়ে 60-বার চেলে দেখেছ। তাতে দেখা গেছে যে, 1 এসেছে 6-বার, 2 এসেছে 10-বার, ইত্যাদি যেমন যেমন দিয়েছে। যদি সব দিকই সমান হত, তবে 60 বারের মধ্যে ঠিক $60 \times \frac{1}{6} = 10$ বার করে প্রতিটা সংখ্যা আসাই প্রত্যাশিত ছিল। সেটা আসেনি। কিন্তু অমনি "ছক্কাটায় নির্ঘাত গোলমাল আছে" বলে মনে করার কোনো কারণ নেই। কারণ হাজার হোক, ব্যাপারটা পুরোই random, একটু এদিক ওদিক তো হতেই পারে। যেমন এখানে 10-এর জায়গায় কখনও 6 এসেছে, কখনো 12 এসেছে। কথাটা হল--এটাকে কি খালি "একটু" এদিক ওদিক বলা যায়, নাকি "অনেকটা" এদিক ওদিক বলা উচিত? ঠিক যেমন দুটো বাচ্চা খেলতে খেলতে একজনের নখে অন্যজনের মুখ ছড়ে গেছে। যার নখে ছড়েছে তার মা বলছে, "আহা, অমন তো একটু আধটু হতেই পারে!"। আর যার মুখ ছড়েছে তার মা চেঁচিয়ে পাড়া মাথায় করে তুলেছে--"আমার সোনার বাছার

মুখের মাংস খুবলে তুলে নিয়েছে গো!" এরকম দুই মায়ের ঝগড়া শুনলে মনে হয় একটা নিরপেক্ষ সিদ্ধান্ত করার পথ থাকলে সুবিধা হত। ঠিক সেরকমই একটা নিরপেক্ষ সিদ্ধান্ত করতে বলেছে এই অংকটায়।
আমরা এখন জজসাহেব হয়ে বিচার করতে বসেছি। প্রথমেই লিখে নেওয়া দরকার দুই পক্ষে কে কে আছে--

> To test
>
> H_0 : the die is unbiased
>
> against
>
> H_1 : the die is biased.

এবার কিছু নাম দিয়ে নিই--

> For each possible outcome $i \in \{1,2,3,4,5,6\}$, we define
>
> - n_i = observed frequency
> - m_i = expected frequency under H_0.

আমাদের এখানে 6-খানা outcome সম্ভব, $1,2,3,4,5$ আর 6. সেই মত আমরা $n_1,...,n_6$ নিয়েছি। এখানে observed frequency মানে "কতবার এসেছে"। যেমন n_1 হল "কতবার 1 এসেছে"। আমাদের অংকে $n_1 = 6$. একইভাবে $n_2 = 10, n_3 = 8$ ইত্যাদি। এদের সঙ্গে তুলনা করে দেখব expected frequency-গুলো অর্থাৎ "কতবার করে আসা উচিত ছিল"। আমরা এদের নাম দিয়েছি $m_1,...,m_6$. যদি ছক্কার সব পাশই সমান হয়ে থাকে (অর্থাৎ যদি H_0 ঠিক হয়), তবে প্রত্যেকটা পাশই মোটামুটি $60 \times \frac{1}{6} = 10$ বার করে আসা প্রত্যাশিত, মানে--

> Then
> $$m_i = 60 \times \frac{1}{6} = 10 \text{ for } i = 1,...,6.$$

এবার দেখতে হবে $n_1,...,n_6$-গুলো $m_1,...,m_6$-এর থেকে কতটা দূরে। সেই দূরত্বটা মাপার একটা বিশেষ ফর্মুলা আছে, তাকে বলে **frequency χ^2-statistic.** জিনিসটা এইরকম--

> Then the frequency χ^2-statistic is
> $$\chi^2 = \sum_1^6 \frac{(n_i - m_i)^2}{m_i} = \cdots = 3.4.$$

এই ফর্মুলাটা কোথা থেকে আসে, সে গল্পটা একটু কঠিন, তার মধ্যে যাচ্ছি না। কিন্তু ফর্মুলাটা যে খুব বিদঘুটে কিছু নয়, সেটা বুঝে নেওয়া যাক। যদি বলতাম দুটো বিন্দু (a_1,a_2) আর (b_1,b_2)-র মধ্যে দূরত্ব বার করতে, তবে তুমি লিখতে $\sqrt{(a_1-b_1)^2 + (a_2-b_2)^2}$. এই square root-টা নিয়ে কাজ করা ঝামেলা। তাই আমরা অনেক সময়ে দূরত্বটার square নিয়ে কাজ করি, মানে $(a_1-b_1)^2 + (a_2-b_2)^2$ নিয়ে। একইভাবে (a_1,a_2,a_3) আর (b_1,b_2,b_3)-এর মধ্যে দূরত্বের square হবে $(a_1-b_1)^2 + (a_2-b_2)^2 + (a_3-b_3)^2$. সেই হিসেবে $(n_1,...,n_6)$ আর $(m_1,...,m_6)$-এর মধ্যে দূরত্বের square হওয়া উচিত
$$(n_1 - m_1)^2 + \cdots + (n_6 - m_6)^2.$$
প্রথম term-টা বলছে n_1 আর m_1-এর মধ্যে দূরত্ব কত, দ্বিতীয়টা বলছে n_2 আর m_2-র মধ্যে, এইরকম। আমাদের অংকে $n_1 = 6$ আর $m_1 = 10$. যাদের পার্থক্য 4. যদি ধর আমরা 60-বারের বদলে 6000 বার ছক্কা চালতাম, তাহলে

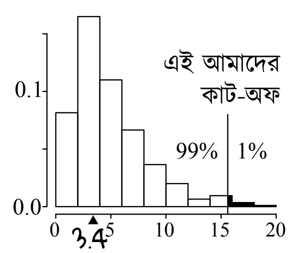


Fig 1

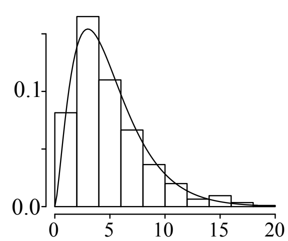


Fig 2

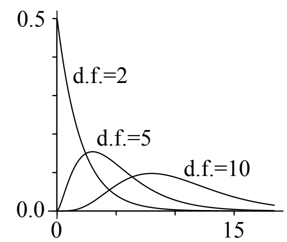


Fig 3

$m_1 = 6000 \times \frac{1}{6} = 1000$ হত। তখনও যদি $m_1 - n_1 = 4$ হত, তাহলেও কী সেই পার্থক্যটা একইরকম গুরুত্বপূর্ণ থাকত? না, কারণ 1000-এর মধ্যে 4 এদিক ওদিক মানে সামান্যই ভুল, কিন্তু 10-এর মধ্যে 4 এদিক ওদিক মানে অনেক বেশী ভুল। এখানেও সেই absolute আর relative-এর গল্প। সেই কারণে আমরা $(m_i - n_i)^2$-গুলোকে m_i দিয়ে ভাগ করে নেব। তাহলেই পাবে frequency χ^2-statistic-এর ফর্মুলাটা--

$$\frac{(n_1 - m_1)^2}{m_1} + \cdots + \frac{(n_6 - m_6)^2}{m_6}.$$

আশা করি বুঝতে পারছ যে, এই সংখ্যাটা যত বড় হবে ততই আমাদের সন্দেহ বাড়বে যে ছক্কাটায় কিছু গোলমাল আছে। Statistics-এর ভাষায়--

> We shall reject H_0 if the observed value of χ^2 is too large.

প্রশ্ন হল "too large" মানে কী। ঠিক কতটা বড় হলে আমরা H_0-কে reject করব? যদি H_0-টা ঠিকও হয়, তাহলেও তো আর এই frequency χ^2 statistic-টা একেবারে 0 হবেই এমন কথা নেই। হাজার হোক, এটা একটা random variable, তাই 0-র থেকে একটু বেশী হতেই পারে। আমরা যে data থেকে 3.4 পেয়েছি, সেটা কী 0-র চেয়ে "একটু বেশী" নাকি "অনেকটা বেশী"? এই সমস্যার সুরাহা করা জন্য একটা কৌশল লাগবে যেটা বুঝতে বহু ছাত্রের প্রথম প্রথম অসুবিধা হয়। তাই এই জায়গাটা ভালো করে বুঝে বুঝে পড়। একটা উদাহরণ নিই--

> সৌরভ ক্লাস টুয়েলভ-এ 93.2% পেয়েছিল। একে কী "খুব বেশী" নম্বর বলা চলে? উত্তরটা নির্ভর করছে কোন বোর্ড তার উপরে। যদি West Bengal Board হয়, তবে অবশ্যই একে "খুব বেশী" বলতে আপত্তি থাকার কারণ নেই। কিন্তু যদি ISC পরীক্ষা হয়, তবে 93.2% তেমন কিছু না। কারণ সেখানে এর চেয়েও অনেক বেশী বেশী নম্বর ওঠে।
>
> এই কারণেই আজকাল percentage না বলে percentile বলার চল হয়েছে। যদি বলি একজনের percentile হল 99, তার মানে সেই বোর্ডের পরীক্ষায় 99% ছাত্র ওর চেয়ে কম বা সমান নম্বর পেয়েছে। সুতরাং আমরা কোনো নম্বর বেশী নাকি কম সেই সিদ্ধান্ত করার জন্য সেই বোর্ডের 99-th percentile-কে কাট-অফ ধরতে পারি (ধরো c) সেই বোর্ডের কোনো ছাত্র যদি $> c$ পায়, তবেই খালি বলব যে, "খুব বেশী" নম্বর পেয়েছে। এখানে তোমার রুচি অনুযায়ী তুমি 99-এর জায়গায় 95 বা 90-th percentile-ও নিতে পারো।

সুতরাং লক্ষ কর যে, একজন ছাত্রের নম্বর "খুব বেশী" কিনা সেই সিদ্ধান্ত করার জন্য যাবতীয় ছাত্রদের নম্বরের distribution জানার দরকার পড়ল। একই যুক্তি আমাদের test-এর বেলায় লাগাতে গেলে জানা দরকার H_0 যদি ঠিক হয়, তবে frequency χ^2 statistic-এর distribution-টা কী হবে। একে বলে null distribution (যেহেতু null hypothesis-কে ঠিক ধরে নিয়ে বার করা distribution)। অর্থাৎ যদি ছক্কাটা সত্যিই unbiased হয়, তাহলে frequency χ^2 statistic-টা কীরকম আচরণ করত। যদি দেখি আমাদের data থেকে পাওয়া value-টা তাদের তুলনায় অনেক বড়, তাহলে সন্দেহ করব যে ছক্কাটা বোধহয় unbiased নয়। Null distribution-টা বার করার একটা কায়দা হল সত্যি সত্যি একটা unbiased ছক্কা নিয়ে বার বার চেলে frequency χ^2 statistic-এর value বার করে দেখা। যদি এইভাবে অনেকগুলো χ^2-এর value বার করি (ধরো 1000-টা) তাহলে তাদের ব্যবহার করে একটা histogram পাব। এরকম একটা histogram দেখিয়েছি

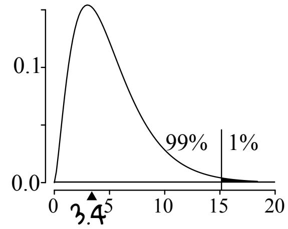


Fig 4

Fig 1-এ। বুঝতেই পারছ যে, χ^2-এর একেকটা value বার করতেই 60 বার করে ছক্কা চালতে হচ্ছে, সুতরাং 1000-টা value-র জন্য মোট $60 \times 1000 = 60000$ বার ছক্কা চালতে হয়েছে। তাই কাজটা খালি হাতে সম্ভব নয়, আমরা কাজটা করেছি কম্পিউটারে, ব্যাপারটা বোঝানোর জন্য। লক্ষ কর যে, এই histogram-টার সাপেক্ষে আমাদের data থেকে পাওয়া value-টা (মানে 3.4) কিন্তু মোটেই "খুব বড়" নয়। যেমন, যদি আমরা 99-th percentile-এ কাট-অফ ধরি, তবে আমাদের 3.4 রয়েছে তার বাঁদিকে। সুতরাং এক্ষেত্রে আমরা H_0-কে accept করব। সমস্যা হল এই যে, এই সিদ্ধান্তে পৌঁছোনোর জন্য আমাদের histogram-টা লাগল, আর সেটা কম্পিউটার ছাড়া আঁকা কঠিন। এবার একটা approximate কায়দা বলব যেটা কম্পিউটার ছাড়াই করা যায়। কায়দাটা হল histogram-টাকে একটা PDF দিয়ে approximate করা, যেমনটা দেখিয়েছি Fig 2-এ। দেখতেই পাচ্ছ approximation-টা বেশ ভালোই মানিয়েছে। যে probability distribution-এর এটা p.d.f., এটা কিন্তু আমাদের পূর্বপরিচিত সেই χ^2-distribution. আমরা এই বইতে আগেই বলেছি যে, এই distribution-এর একটাই parameter, নাম degrees of freedom বা সংক্ষেপে d.f. এর বিভিন্ন value-তে pdf-টা কীরকম দেখতে হয় তার ছবি রয়েছে Fig 3-এ।

এইরকম অংকে d.f. বার করার একটা কায়দা আছে-- d.f.-টা হবে sample space-এর সাইজের থেকে 1 কম। এখানে ছক্কা চালা হয়েছিল, তাই sample space ছিল $\{1, 2, 3, 4, 5, 6\}$, যার সাইজ 6. সুতরাং d.f. হচ্ছে $6 - 1 = 5$. কেন এই কায়দাটা কাজ করে সেটা বোঝানো এই বইয়ের পাল্লার বাইরে।

> We know that if H_0 is true, then the frequency χ^2 statistic is distributed approximately as χ^2 with $6 - 1 = 5$ degrees of freedom,

এই যে approximation-টা, এটা এক ধরণের statistical regularity-র প্রয়োগ। এবং statistical regularity যেখানেই লাগাচ্ছি, সেখানে sample size-টা যথেষ্ট বড় হওয়া চাই। আমাদের ক্ষেত্রে তাই প্রয়োজন--

- মোট sample size (মানে 60) যেন যথেষ্ট বড় হয় (সাধারণতঃ ≥ 30 হলেই কাজ চলে),
- আর প্রতিটা expected frequency-ও যেন যথেষ্ট বড় হয় (মোটামুটি ≥ 5 হলেই চলে)।

আমাদের বেলায় দুটো শর্তই খাটছে--

> because the sample size is ≥ 30 and all the expected frequencies are ≥ 5.

এবার একটা কাট-অফ বার করতে হবে (statistics-এর ভাষায় যাকে বলে critical value), যার উপরে গেলে আমাদের χ^2-এর value-টাকে "too large" আখ্যা দেওয়া যাবে। Critical value-টা নেব ধরো 99-th percentile, মানে সেই সংখ্যাটা যার ডানদিকের area-টা 0.01, যেমনটা দেখিয়েছি Fig 4-এ। সেই সংখ্যাটা এই অংকে বলাই আছে, 15.09. বস্তুতঃ সেটা দেওয়া আছে দেখেই 99-th percentile নিয়েছি। সুতরাং--

> We shall reject H_0 if the observed value of χ^2 exceeds 15.09.

এইবার তবে দেখতে হবে যে, আমরা χ^2-এর যে value পেয়েছিলাম (মানে 3.4), সেটা এর চেয়ে বড় না ছোটো।

> In this case $3.4 \leq 15.09$. So we accept H_0 at 1% level of significance to conclude that the given data set does not provide any evidence against the null hypothesis that the die is unbiased.

এই লেখাটার মধ্যে একটা সূক্ষ্ম আইনের প্যাঁচ আছে। আমরা কিন্তু বললাম না যে "Hence H_0 is true." গতকালের সেই জল-ভরা কৌটোর উদাহরণটা ভাবলে ব্যাপারটা বুঝতে পারবে। কৌটো ঝাঁকিয়ে যে আওয়াজটা পেয়েছি সেটা জলের আওয়াজের মত হলেই অমনি জোর দিয়ে বলা চলে না যে, কৌটোর মধ্যে জলই আছে। তাই আমরা খালি বলতে পারি যে, "জল নেই এমন মনে করার কোনো কারণ নেই"। কিন্তু যদি আমাদের χ^2-টা 15.04 হত, তবে জোর দিয়ে বলতাম যে, 1% level of significance-এর সিদ্ধান্ত করছি ছক্কাটা unbiased নয়। ■

এই যে অংকটা করলাম সেটা ছিল χ^2 test for goodness of fit-এর একটা প্রয়োগ। এখানে আমরা একটা ছক্কার সব পাশ সমান ভারী কিনা সেটা পরীক্ষা করে দেখলাম। ছক্কা ছাড়া অন্য কিছুর ক্ষেত্রেও একই কাজ করা যায়। সেই general রূপটা আমরা শিখব কিছু পরে। কিন্তু এই উদাহরণ থেকেই আমরা যেকোনো statistical test-এরই মূল চেহারাটার সঙ্গে পরিচিত হয়েছি। তার অঙ্গপ্রত্যঙ্গগুলো চিনে রাখা যাক--

1. প্রথমেই আসে দুটো hypothesis, যাদের নাম দেওয়া হয় H_0 আর H_1.

2. তারপর data-র উপর একটা ফর্মুলা লাগিয়ে বার করেছিলাম frequency χ^2 statistic. আমাদের ক্ষেত্রে এটাই ছিল test statistic. অন্য test-এর বেলায় test statistic-টা অন্য কিছু হবে। Test statistic-গুলো কোথা থেকে আসে তা নিয়ে আমরা খুব কিছু বলি নি। সেটা কালকে শিখব।

3. এরপর বুঝতে হয় কী হলে H_0-টা reject করব--test statistic-এর value-টা "খুব বেশী বড়" হলে, নাকি "খুব বেশী ছোটো" হলে। এটা অনেকটা স্বাস্থ্যবিধির মত ব্যাপার। কোলেস্টেরল বেড়ে গেলে খারাপ, আবার লোহিত কণিকার সংখ্যা কমে গেলে খারাপ। কোনো কোনো ক্ষেত্রে আবার বাড়া-কমা দুইই খারাপ, যেমন রক্তচাপ। যদি "খুব বেশী বড়" হলে H_0-কে reject করা উচিত হয়, তবে সেরকম test-কে বলব upper tailed test. উল্টোদিকটাকে বলব lower tailed test. আর রক্তচাপের মত কেস হলে বলব two tailed test. এই কটার বাইরে আর কেস হয় না, তাই হুঁকোমুখো হ্যাংলার মত statistical test-রাও বলতে পারে "দুটি বৈ ল্যাজ মোর নাই রে।" । আমাদের χ^2-test-টা ছিল একটা upper tailed test.

4. এর পরের কাজ ছিল critical value বার করা। তার জন্য প্রথমে দেখতে হয় test statistic-এর null distribution, অর্থাৎ যদি H_0 ঠিক হত, তবে test statistic-এর আচরণ কীরকম হত। এর সঙ্গেই তুলনা করব আমাদের data থেকে পাওয়া test statistic-এর value-টার। এই critical value বার করা হয় percentile-এর কায়দায়। এইখানেই level of significance-টা কাজে লাগে। ধরো level of significance হল 1%. তাহলে এমনভাবে critical value নেওয়া হয়, যাতে null distribution-এর 1% অংশই খালি critical value-র বাইরে পড়ে।

5. সবশেষে সিদ্ধান্তটাকে লিখতে হয় আইনের সূক্ষ্ম প্যাঁচটা রেখে, মানে H_0-কে accept করতে হলে বলা হয় যে, H_0-এর বিপক্ষে এই data থেকে যথেষ্ট সাক্ষ্য পাওয়া যাচ্ছে না।

একটা উদাহরণের থেকেই অনেক কিছু বলা গেল। সবকিছু হজম হতে একটু সময় নেবে। চিন্তা নেই, এই একই প্যাটার্ণ যেকোনো statistical test-এই থাকে। সুতরাং পরপর একই জিনিসের বিভিন্ন উদাহরণ দেখতে দেখতে ব্যাপারটা মাথায় ঢুকে যাবে।

Goodness-of-fit-এর আরেকটা অংক করি, এবার কয়েন টসের। এর জন্য একটা সহজ তথ্য মনে করিয়ে দিই--যদি একটা কয়েনকে n বার টস করে k-খানা head পাওয়া যায়, তবে $P(\text{head})$-এর MLE হয় $\frac{k}{n}$.

Example 15: আমরা যখন প্রথম অধ্যায়ে statistical regularity-র আলোচনা করেছিলাম, তখন একই random

experiment বহুবার করার কথা আসছিল। একজন মাষ্টারমশাই ছাত্রদের সেটা হাতেকলমে করিয়ে দেখাতে চান, তাই তিনি ঠিক করেছেন যে, ছাত্রদের দিয়ে কোনো একটা কয়েনকে 400 বার টস করিয়ে কতবার head পড়ে গোণাবেন। ক্লাসে 100

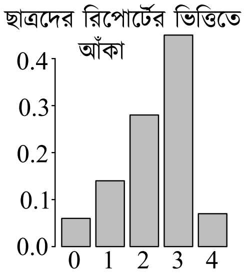


Fig 5

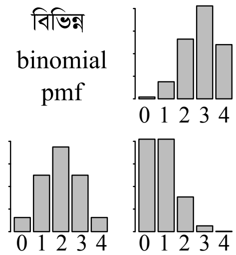


Fig 6

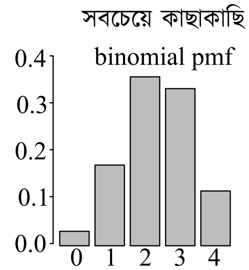


Fig 7

জন ছাত্র আছে, তিনি তাই নির্দেশ দিলেন যে, তোমরা হোস্টেলে গিয়ে কোনো একটা কয়েন নেবে, তারপর প্রত্যেকে সেটাকে 4 বার করে টস করে পর দিন আমাকে বলবে কে কটা head পেলে। সেই মত পর দিন 100 জন ছাত্র স্যারের কাছে বিভিন্ন সংখ্যা রিপোর্ট করল। আমরা তাদের frequency distribution-টা দিলাম--

No. of heads	Frequency
0	6
1	14
2	28
3	45
4	7
Total	100

বুঝতেই পারছ যে, সবগুলো সংখ্যাই $\{0, 1, 2, 3, 4\}$-এর মধ্যে, কারণ 4 বার টস করলে head-এর সংখ্যা এর বাইরে কিছু হতে পারে না। এখন মাষ্টারমশাই গোড়াতেই যাচাই করে নিতে চান ছাত্ররা সবাই ঠিকঠাকভাবে একই কয়েন টস করেছে কিনা। কীভাবে করবেন?

SOLUTION: তিনি প্রথমেই একটা barplot এঁকে নিলেন, যেমন দেখিয়েছি Fig 5-এ। দেখেই ওনার একটু সন্দেহ হল। যেহেতু একই কয়েনকে 4 বার টস করে head-এর সংখ্যা গোণার কথা তাই এই barplot-এর চেহারাটা কোনো একটা Binomial$(4, p)$-এর p.m.f.-এর মত হওয়া উচিত ছিল। অবশ্য p-টা মাষ্টারমশায় জানেন না, কারণ হোস্টেলে যে কয়েনটা টস করা হয়েছিল সেটা তাঁর কাছে জমা দেওয়া হয় নি। কিন্তু তাও binomial family-র p.m.f.-এর আদলটা মাষ্টারমশায়ের অজানা নয়, সেটা Fig 6-এর মত কিছু একটা দেখতে হয়। অবশ্য এ কথা ঠিক যে, data হল random, তাই তার ভিত্তিতে পাওয়া barplot যে একেবারে নিঁখুতভাবে p.m.f.-এর সাথে মিলবে এমন কোনো কথা নেই, কিন্তু statistical regularity-র কথাটাও ভুললে চলবে না। "4 বার টস করা"-কে যদি একটা random experiment ধরি, তবে সেটা 100 বার করা হয়েছে, সেটা খুব কম নয়। সুতরাং data থেকে আঁকা barplot-টা p.m.f.-এর থেকে খুব দূরে থাকা উচিত নয়। কিন্তু এখানে তো একেবারেই সেরকম মনে হচ্ছে না! সন্দেহটাকে অংকের ভাষায় গুছিয়ে লেখা যাক। এই কাজটা অনেকটা ছক্কার অংকটার মতই, data-র barplot-এর সঙ্গে একটা p.m.f.-এর তুলনা করা। খালি এখানে সমস্যা হল তুলনাটা ঠিক কোনো একটা p.m.f.-এর সঙ্গে হচ্ছে না, যেহেতু $p = P(\text{head})$-টা মাষ্টারমশায়ের অজানা। তাই তুলনাটা করতে হচ্ছে গোটা binomial family-র যাবতীয় p.m.f.-এর আদলের সঙ্গে। মাষ্টারমশায় p জানেন না বটে, কিন্তু সেটা উনি estimate করতে পারেন। তার জন্য প্রথমে গুণে দেখলেন মোট কতগুলো head পাওয়া গেছে--

$$0 \times 6 + 1 \times 14 + 2 \times 28 + 3 \times 45 + 4 \times 7 = 233.$$

সুতরাং ছাত্ররা যদি সবকিছু ঠিকঠাক করে থাকে, তবে 400-টা টসে 233-টা head পাওয়া মানে p-এর MLE হবে $\widehat{p} = \frac{233}{400} = 0.5825$. অতএব তিনি ছাত্রদের থেকে পাওয়া frequency distribution-এর সঙ্গে Binomial$(4, \widehat{p})$-এর তুলনা করতে বসলেন। এইটা হল Binomial$(4, p)$ family-র সেই সদস্য যেটা data-র সবচেয়ে কাছে আছে। সুতরাং আমাদের data যদি এই family-র কারুর সঙ্গে ফিট করে তো সেটা এর সঙ্গেই করবে। Fig 7 দ্যাখো। এবার যেহেতু p-এর জায়গায় একটা সংখ্যা বসিয়ে নিয়েছেন, তাই ছক্কার উদাহরণের মত χ^2-test-এর পথে এগোনো যেতেই পারে। সুতরাং নীচের টেবিলটা পাওয়া গেল।

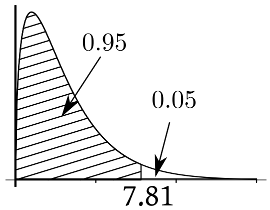


Fig 8

k	n_i	$\widehat{p}_i$	$\widehat{m}_i = n\widehat{p}_i$	$(n_i - \widehat{m}_i)^2/\widehat{m}_i$
0	6	0.03038267	12.15307	3.115283
1	14	0.1695608	67.8243	42.71412
2	28	0.3548592	141.9437	91.467
3	45	0.3300686	132.0275	57.36517
4	7	0.1151287	46.05149	33.11552

এখানে

$$\widehat{p}_i = \binom{4}{i}\widehat{p}^i(1-\widehat{p})^{4-i}, \quad i = 0, 1, 2, 3, 4.$$

শেষের column-টা যোগ করে পাচ্ছি 227.78. প্রশ্ন হল এটা খুব বড় কিনা। তার জন্য critical value বার করব যথারীতি χ^2 distribution ব্যবহার করে। এই অব্দি ছক্কার উদাহরণের মতই। পার্থক্যটা হবে degrees of freedom বার করার সময়ে। আগের উদাহরণে আমরা sample space-এর সাইজের থেকে 1 বিয়োগ করেছিলাম। এখানেও তাই করব, কিন্তু তার থেকে আরও 1 বাদ দেব, যেহেতু একটা parameter আমরা estimate করেছি। সুতরাং degrees of freedom হবে $5-1-1=3$. আমরা যদি 5% level of significance ব্যবহার করি তবে critical value-টা হবে 7.81. এই value-টা কী করে এল সেটা বোঝার জন্য Fig 8 দেখলে সুবিধা হবে। এইরকম critical value কম্পিউটার ব্যবহার করে PDF-কে integrate করে বার করা যায়, বা χ^2-distribution-এর টেবিল থেকে পাওয়া যায়। যেহেতু $227.78 > 7.81$, তাই আমরা রায় দেব-- data দেখে মনে হচ্ছে না যে, ছাত্ররা একটাই কয়েনকে প্রত্যেকে 4 বার টস করেছে। ■

এই উদাহরণ থেকে শিখলাম অজানা parameter-এর উপস্থিতিতে কীভাবে χ^2 goodness-of-fit test করতে হয়। এবার আরেকটা উদাহরণ দেখি, সেখানে আরেকটা নতুন জিনিস আসবে।

Example 16: সজ্জন সচ্চরিত্র ছক্কাদের সবগুলো পাশই একরকম ভারী হয়। কিন্তু মানুষদের মধ্যে যেমন সবাই সজ্জন সচ্চরিত্র হয় না, ছক্কাদের বেলাতেও তাই। জুয়ার আসরে গোলমেলে ছক্কাদের কদর অনেক। কথাশিল্পী পরশুরাম তাঁর হাসির গল্প "তৃতীয় দ্যূতসভা"-তে এই রকম একটা ছক্কার কথা লিখেছিলেন, তার মধ্যে এটা আরশোলা ভরা থাকত, তাই সহজে উল্টোতো না। অবশ্য এত করেও খুব একটা সুবিধা হয় নি। কারণ আরেকজনের ছক্কার কাছে এই ছক্কাও বেকুব বনে গিয়েছিল (সেই ছক্কাটাতে একটা আস্ত টিকটিকি ভরা ছিল)! যাই হোক, এসবই নিছক গল্প। বাস্তবে জালি ছক্কা তৈরীর প্রধানতঃ দুটো কায়দা। এক হল ছক্কাটার ভিতরে এক পাশে একটা লোহার বা সীসার টুকরো ঢুকিয়ে দেওয়া। তাতে সেই পাশটা নীচের দিকে থাকার প্রবণতা বেড়ে যাবে। তবে এরকম ছক্কা তৈরী করা কঠিন। এর চেয়ে সহজ হল এমনভাবে বানানো যাতে ছক্কাটা পুরো cube না হয়ে একটু অসমান হয়, অনেকটা ইঁটের মত। এরকম ছক্কার ক্ষেত্রে পরস্পর বিপরীত দুটো পিঠের probability সমান হবে, যেমন 1 আর 6-এর, আবার 2 আর 5-এর, এবং 3 আর 4-এর। কিন্তু 1-এর probability আর 2-এর probability সমান নাই হতে পারে। এবার ধরো একটা ছক্কা বারবার চালা হয়েছে। তার outcome-গুলো হয়েছে এইরকম--

i	1	2	3	4	5	6
n_i	55	101	141	163	86	54

এই data-র ভিত্তিতে তোমাকে পরীক্ষা করতে বলা হল যে ছক্কাটা এরকম একটা cuboid-এর মত কিনা। কী করে করবে?
SOLUTION: যদি আমরা i পড়ার probability-টাকে p_i নাম দিই, তবে আমাদের দেখতে হবে

$$H_0 : p_1 = p_6 \text{ and } p_2 = p_5 \text{ and } p_3 = p_4$$

কিনা। লক্ষ কর যে এখানে p_i-গুলোকে বার করে ফেলা যাচ্ছে না। কিন্তু যদি $p_1 = p_6 = \theta_1$, $p_2 = p_5 = \theta_2$ আর $p_3 = p_4 = \theta_3$ নাম দিয়ে নিই, তবে ছয়টা p_i-কে তিনটে θ_i দিয়ে প্রকাশ করা যাচ্ছে। কিন্তু এখনও যেহেতু θ_i তিনটে অজানা, তাই expected frequency-গুলো বার করা যাচ্ছে না। তবে উপায়? উপায় হল θ_i-গুলোকে estimate করা, এবং এই কাজে আমরা MLE ব্যবহার করব। MLE বার করার জন্য অবশ্যই likelihood-টা পুরো লিখে maximise করা উচিত। কিন্তু আমরা চট্‌ করে বোঝানোর জন্য সংক্ষেপে এগোব এইভাবে--

> লক্ষ কর যে θ_1 হল 1 আর 6 পড়ার probability. এখানে 1 আর 6 মোট এসেছে $55 + 54 = 109$ বার। তাই 1 আর 6 পড়ার probability আমরা estimate করতে পারি $\frac{109}{600} = 0.18$. যেহেতু 1 আর 6 পড়ার probability সমান, তাই ওদের প্রত্যেকের ভাগে পড়বে এর অর্ধেক করে। মানে θ_1-এর MLE হবে $\widehat{\theta}_1 = 0.18/2 = 0.09$.

একইভাবে 2 আর 5 মিলিয়ে $\widehat{\theta}_2 = 0.16$. এভাবে $\widehat{\theta}_3$-ও বার করা যাবে, কিন্তু তার আর দরকার নেই। কারণ লক্ষ কর যে $\theta_1 + \theta_2 + \theta_3 = \frac{1}{2}$ হতে বাধ্য। সুতরাং বিয়োগ করেই পেয়ে যাচ্ছি

$$\widehat{\theta}_3 = \tfrac{1}{2} - \widehat{\theta}_1 - \widehat{\theta}_2 = 0.25.$$

সুতরাং H_0 ঠিক হলে আমরা নীচের টেবিলটা পাচ্ছি--

i	n_i	$\widehat{p}_i$	$\widehat{m}_i = n\widehat{p}_i$
1	55	0.09	54.5
2	101	0.16	93.5
3	141	0.25	152
4	163	0.25	152
5	86	0.16	93.5
6	54	0.09	54.5

এখানে χ^2-statistic-টা হবে

$$\sum \frac{(n_i - \widehat{m}_i)^2}{m_i} = 2.80.$$

প্রশ্ন হল--এটা কী "যথেষ্ট বড়"? এর জন্য জানা দরকার H_0 ঠিক হলে এই statistic-টার আচরণ (মানে distribution) কীরকম হয়। উত্তর হল χ^2 distribution, যেখানে degrees of freedom হল

> sample space-এর সাইজ $-1-$ no. of estimated free parameters.

এখানে sample space-এর সাইজ হল 6. আমাদের তিনটে parameter ছিল θ_1, θ_2 আর θ_3. কিন্তু এদের মধ্যে একটা সম্পর্ক ছিল $\theta_1 + \theta_2 + \theta_3 = \frac{1}{2}$. তাই যে কোনো দুটো estimate করলেই, বাকিটা বিয়োগ করে বার করে দেওয়া যাচ্ছিল। সুতরাং "no. of estimated free parameters" হল 2. অতএব সব মিলিয়ে degrees of freedom দাঁড়াচ্ছে $6 - 1 - 2 = 3$.
এর জন্য 5% level-এ critical value হল 7.81. যেহেতু $2.8 < 7.81$,তাই আমরা এখানে H_0-টা accept করব। ■

এতক্ষণ আমরা যা যা শিখলাম, সেগুলোকে এবার গুছিয়ে লেখার সময় এসেছে নীচের অংকের উত্তরে--

Example 17: Write a short note on χ^2-test of goodness-of-fit of a random sample to a hypothetical distribution.[4] **(2005.10a)**
SOLUTION:

> Let $X_1, ..., X_n$ be iid from a discrete distribution with k categories, $1, ..., k$.
> Let $p_1(\theta), ..., p_k(\theta)$ be a hypothetical distribution, where $\theta \in \Theta$ is unknown, but p_i is a known function of θ.

যেমন আগের উদাহরণে আমরা 6-টা p_i-কে 3-টে θ_i দিয়ে প্রকাশ করতে পেরেছিলাম।

> We want to test
>
> H_0 : For each $i \in \{1, ..., k\}$ the probability of category i is $p_i(\theta)$ for some $\theta \in \Theta$.
>
> Based on the data we compute MLE $\widehat{\theta}$ of θ assuming H_0 to be true. Then, for each category, we estimate the expected frequency under H_0 as $\widehat{m}_i = np_i(\widehat{\theta})$.
> Then we compute
>
> $$\text{frequency } \chi^2 \text{ statistic} = \sum \frac{(n_i - m_i)^2}{m_i}.$$
>
> We assume that n is large (say $n \geq 30$) and each $\widehat{m}_i$ is large (say, $\widehat{m}_i \geq 5$). Then under H_0 the frequency χ^2 statistic approximately follows χ^2 distribution with degrees of freedom given by
>
> $k - 1 -$no. of estimated free parameters.
>
> Let F denote the cdf of this distribution. Then we reject H_0 at α level of significance if and only if $T > F^{-1}(1 - \alpha)$.

■

Example 18: Of 160 offsprings of a certain a cross between guineapigs, 102 were found to be red, 24 black, 34 white. According to a genetic model the probabilities of red, black and white are $\frac{9}{16}, \frac{3}{16}, \frac{1}{4}$, respectively. Test at 5% significance level if the data are consistent with the model. [Given that $\chi^2_{0.05} = 5.99$ for 2 d.f.□] **(2008.5c)**

SOLUTION: একেবারে প্রথমে আমরা ছক্কার যে অংকটা করেছিলাম, এই অংকটা ঠিক সেটারই মত। ওখানে random experiment-টা ছিল ছক্কা চালা, আর এখানে গিনিপিগ জন্মানো। ছক্কা চালার পর আমরা যেমন outcome হিসেবে দেখি কোন সংখ্যাটা পড়েছে, তেমনি একটা গিনিপিগের ছানা জন্মালে এখানে আমরা ব্যস্ত হয়ে দেখছি তার গায়ের রং কী হল। তিন রকম রং হতে পারে--লাল, কালো আর সাদা। রংটা নির্ভর করে বাবা আর মায়ের কী রং ছিল তার উপরে। ঠিক কীভাবে এই রংটা এক প্রজন্ম থেকে পরবর্তী প্রজন্মে যায়, তা নিয়ে গবেষণার আর অন্ত নেই। পণ্ডিতেরা নানারকম তত্ত্ব খাড়া করেন সে নিয়ে। একটা করে তত্ত্ব খাড়া হয়, আর তারপর ল্যাবরেটরীতে গিনিপিগদের ছানা জন্মিয়ে তাদের গায়ের রং দেখা হয়। যদি সেই তথ্যের সঙ্গে তত্ত্ব মিলল তো ভালো, নইলে সেই তত্ত্ব বাতিল করে আবার নতুন তত্ত্ব খাড়া করতে হয়। এখানে আমাদেরকে এই তথ্যের সঙ্গে তত্ত্ব মেলানোর কাজটাই করতে বলেছে।

Here each random offspring is like a trial of a random experiment. The colour of the offspring can be Red, Black or White.

Let

- n_i = observed frequency of colour i,
- p_i = postulated probability of colour i,
- m_i = expected frequency of colour i.

We want to test

H_0 : the given data are consistent with the postulated probabilities

against

H_1 : Not H_0.

Based on the given data we construct the following table.

i	Colour	n_i	p_i	$m_i = np_i$
1	Red	102	$\frac{9}{16}$	90
2	Black	24	$\frac{3}{16}$	30
3	White	34	$\frac{1}{4}$	40
Total		160(= n)	1	160

So the goodness-of-fit χ^2-statistic is

$$\sum_1^3 \frac{(n_i - m_i)^2}{m_i} = 3.7.$$

Here the total sample size $n = 160 \geq 30$ and each $m_i \geq 5$.
So under H_0 this statistic has approximate χ^2-distribution with $3 - 1 = 2$ degrees of freedom. So we reject H_0 at 5% level of significance if the observed value of the statistic exceeds the upper 5% quantile of the χ^2-distribution with 2 degrees of freedom, which is given to be 5.99.

Since $3.7 \leq 5.99$, hence we accept H_0 to conclude that, at 5% level of significance, we do not have significant evidence against believing that the data are consistent with the model.

■

DAY 37
Neyman-Pearson technique

গত পরশু আমরা শিখেছি test of hypothesis কী এবং তার অঙ্গপ্রত্যঙ্গের নাম কী কী। গতকাল শিখেছি একটা বিশেষ ধরণের test-এর কথা, যার নাম χ^2 goodness-of-fit test. এখনও কিন্তু বলি নি কী করে একটা test বানাতে হয়। এবার সেই আলোচনা করব। আমরা test তৈরী করার যে কায়দাটা শিখব সেটা Jerzy Neyman এবং Egon Pearson[1] নামক দুই পণ্ডিতের তৈরী, তাই একে অনেক সময়ে Neyman-Pearson technique বলে। এই কায়দাটার মূল ব্যাপারটা নিতান্তই সহজ--একটা test বানানোর সময়ে এমনভাবে বানানো উচিত যাতে ভুল সিদ্ধান্তের সম্ভাবনা যথাসম্ভব কম থাকে। ভুল সিদ্ধান্ত যে দুই রকম হয় সে তো বলেইছি, type I আর type II. এদের সংজ্ঞাগুলো সংক্ষেপে মনে করিয়ে দিই--

আমাদের রায়

আসলে কোনটা ঠিক		accept H_0	reject H_0
	H_0	✔	type I error
	H_1	type II error	✔

ডাক্তার বলেছেন ...

আসলে ক্যান্সার ...		ক্যান্সার হয়েছে	ক্যান্সার হয়নি
	হয়েছে	✔	type I error
	হয়নি	type II error	✔

বাঁদিকের টেবিলটা থেকে দুই type-এর error-এর সংজ্ঞাগুলো পাচ্ছ, আর ডানদিকেরটায় আমাদের পূর্বপরিচিত ক্যান্সারের উদাহরণ। আমরা দুই ধরণের error-এর probability-ই কমাতে চাই, কিন্তু সমস্যা হল একসাথে দুটোকেই কমানো সাধারণতঃ অসম্ভব। ক্যান্সারের উদাহরণ থেকেই সেটা বুঝবে--

Example 19: মনে কর এক ডাক্তারবাবু আছেন যিনি অতিসাবধানী, তাঁর ভয় পাছে তিনি একজন ক্যান্সারের রোগীকেও ভুল করে ফিরিয়ে দেন (মানে type I error করে বসেন)। সেই ভয়ে তিনি করেন কি, টিউমার দেখলেই বলেন নির্ঘাত ক্যান্সার হয়েছে। এই পদ্ধতিতে type I error-এর probability যে 0, সেটা বলাই বাহুল্য। কিন্তু যাদের ক্যান্সার হয় নি, সে বেচারীদের দুরবস্থা ভাবো। তাদের প্রতি ডাক্তারবাবু সবসময়েই অবিচার করছেন, মানে type II-এর probability হয়ে যাচ্ছে 1. আবার আরেক ডাক্তার আছেন মনে করো, যিনি বলেন ক্যান্সার-ট্যান্সার সব বাজে কথা, আমার দুই পুরিয়া ওষুধ খেলেই টিউমার আপনেই সারবে। ইনি type II error করছেন না একেবারেই, কিন্তু type I error-এর probability এনার ক্ষেত্রে 1. সুতরাং আমাদের হয়েছে মারীচের অবস্থা, এদিকে গেলে রাম মারবে, ওদিকে গেলে রাবণ মারবে। ■

Neyman-Pearson-এর কায়দাটা হল এরই একটা সমাধান, শ্যাম আর কুল দুইই রাখার একটা চেষ্টা। এর জন্য প্রথমেই লক্ষ কর যে, type I error-টা type II-এর চেয়ে বেশী ভয়ানক। তাই Neyman আর Pearson-এর উপদেশ হল P(type I error)-কে কোনোভাবেই বেশী বাড়তে না দেওয়া। প্রথমেই তার জন্য একটা upper bound বেঁধে দাও, ধরো 1%. (এই upper bound-এর একটা নাম যে level of significance, সে তো আমরা গতপরশুই শিখেছি)। যে সব test-এর বেলায় $P(\text{type I error}) > 0.01$, তাদের কথা আমরা মাথাতেই আনব না। বাকি যারা রইল (মানে যে সব test-এর বেলায় $P(\text{type I error}) \leq 0.01$) তাদের মধ্যে কোন test-এর জন্য P(type II error) সবচেয়ে কম হয় দ্যাখো। সেরকম test-ই হল Neyman আর Pearson-এর মতে সবচেয়ে ভালো test.
এখানে একটা নতুন ভাষা শিখে রাখলে সুবিধা হবে। যেহেতু আমরা সবচেয়ে কম P(type II error)-ওয়ালা test খুঁজছি, তাই ঘুরিয়ে বলা যায় যে আমরা সবচেয়ে বেশী power-ওয়ালা test খুঁজছি (কারণ power $= 1 - P(\text{type II error})$)।

[1] আমরা যে χ^2 test-এর কথা শিখলাম গতকাল, সেটাও জনৈক Pearson-এর তৈরী। তিনি ছিলেন Karl Pearson. আর ইনি তাঁরই ছেলে Egon Pearson.

তাই এরকম test-কে বলে **most powerful test,** বা সংক্ষেপে **MP test.** এদিকে Neyman আর Pearson-এর আদ্যাক্ষর জুড়লে হয় NP. সুতরাং কখনো এনাদের কায়দায় পাওয়া test-কে লোকে **NP test**-ও বলে।
তবে এখানে বলে রাখি, Neyman-Pearson-এর কায়দাটা কিন্তু সবসময়ে লাগানো যায় না। যদি H_0 আর H_1 দুজনেই simple hypothesis হয় তবে কোনো অসুবিধা নেই, কিন্তু composite hypothesis-এর বেলায় অসুবিধা হতে পারে। সে সব খুঁটিনাটির মধ্যে এখনই যাব না। যা যা বললাম সেগুলো আগে গুছিয়ে লিখতে শেখা যাক।

Example 20: How are type I and type II errors used to accept or reject a notional hypothesis?[3]

(2013.6b)
SOLUTION: এখানে "notional" শব্দটা দেখে ভয় পেও না। ওর অর্থ হল "যেটা ধরে নেওয়া হয়েছে এমন"। যেমন আমরা আলোচনার সময়ে বলে থাকি "ধরো যদি অমুক ব্যাপারটা হত", সেইরকম। Statistics-এ hypothesis যখন করা হয় সেটা notional-ই হয়, সেটা আলাদা করে বলার দরকার পড়ে না। আসলে এখানে notional hypothesis বলতে null hypothesis বোঝাতে চেয়েছে। Statistics-এর আধুনিক ভাষায় এই অর্থ আর ব্যবহৃত হয় না।

> When designing a test our aim is to reduce the probabilities of both types of errors as much as possible. However, usually $P(\text{type I error})$ can be reduced only by increasing $P(\text{type II error})$, and vice versa. So reducing both of them simultaneously is impossible.
>
> The Neyman-Pearson technique to construct an optimal test is to specify an upper bound α for $P(\text{type I error})$, and then to choose a test that minimises $P(\text{type II error})$ among all tests with $P(\text{type I error}) \leq \alpha$.
>
> Such a test may not exist in general. However, if both the null and alternative hypotheses are simple, then such a test always exists.

■

Example 21: Explain how the best critical region is determined.[2] **(2011,2009,2007)**

SOLUTION: এটা আসলে আগের অংকটাই, কিন্তু ভাষাটা অনেকটাই অন্যরকম। এখানে test কথাটাই বলে নি। মনে রেখো যে, test মানে হল এমন একটা নিয়ম যেটা ব্যবহার করে তুমি data-র ভিত্তিতে বলতে পারবে কখন H_0-কে accept করবে, আর কখনই বা reject করবে। যেহেতু accept আর reject ছাড়া আর কিছু করার নেই, তাই ঠিক কোন কোন data-র জন্য reject করবে জানতে পারলেই হল (তার বাইরে কোনো data এলেই তার জন্য H_0-কে accept করবে)। যে সব data-র জন্য H_0-কে reject করা হয়, তাদের set-কেই বলে critical region, সেটা আমরা গত পরশুই শিখেছিলাম। সুতরাং একটা test বার করা আর তার critical region বার করা একই কথা। আর এখানে best মানে হল Neyman-Pearson-এর মতে যেটা best. সুতরাং মোদ্দা কথাটা দাঁড়ালো "Neyman-Pearson-এর কায়দাটা সংক্ষেপে লেখো।" সেটা আগের অংকের উত্তর থেকেই পেয়ে যাবে। ■

37.1 কী করে বার করব?

Neyman-Pearson-এর কায়দাটার প্রথম ধাপগুলো আমরা শিখলাম এইরকম--

1. প্রথমেই $P(\text{type I error})$-র একটা upper bound (মানে level of significance) ঠিক করো, ধরো α.
2. তারপর দেখতে হবে সেইসব যাবতীয় test যাদের এর বেলায় $P(\text{type I error}) \leq \alpha$ হয়, তাদের মধ্যে থেকে এমন একটা test বার করো যার জন্য $P(\text{type II error})$ সবচেয়ে কম হয় (বা power সবচেয়ে বেশী হয়)।

প্রথম ধাপটা তো খুবই সহজ যা হোক একটা ছোটো সংখ্যা নিলেই হল (সাধারণতঃ লোকে 1% বা 5% কি বড়জোর 10% নিয়ে থাকে)। সমস্যা হল দ্বিতীয় ধাপটা নিয়ে। "যাবতীয় test যাদের জন্য $P(\text{type I error}) \leq \alpha$ হয়" শুনতে সহজ, কিন্তু এরকম যাবতীয় test-দের তালিকাটা তো আর টেলিফোন ডাইরেক্টারী থেকে পাওয়া যাবে না, google থেকেও না। আর যদিও বা সেই বিশাল তালিকা পেলাম, তার ভিতর থেকে সবচেয়ে বেশী power-ওয়ালা test-কে খুঁজে বার করা তো দুঃসাধ্য বলে মনে হচ্ছে! তবে উপায়? যদি H_0 আর H_1 দুজনেই simple hypothesis হয়, তবে এর একটা আশ্চর্য সহজ উপায় আছে, সেটাও ওই Neyman আর Pearson-ই বার করে গেছেন। সেটাকে বলে Neyman-Pearson Theorem, যেটা চেয়েছে নীচের অংকটায়।

Example 22: State Neyman-Pearson's theorem on the best critical region for testing a simple hypothesis against a simple alternative.[2] **(2005.9a)**

SOLUTION:

Neyman-Pearson theorem

Let $X_1, ..., X_n$ be our data set with likelihood $L(\theta)$. Suppose that we want to test

$$H_0 : \theta = \theta_0 \text{ against } H_1 : \theta = \theta_1.$$

at some given level α. Let

$$LR = \frac{L(\theta_0)}{L(\theta_1)}.$$

Consider a test with $P(\text{type I error}) = \alpha$ that

- rejects H_0 if $LR < c$,
- accepts H_0 if $LR > c$,
- may either accept or reject if $LR = c$.

Then this test must have maximum power among all tests of level α.

■

জব্বর জিনিস! দেখলেই গায় কাঁটা দেয়, এবং মাথামুণ্ডু কিস্‌সু বোঝা যায় না। আসলে কিন্তু এটা একেবারেই সহজ বুদ্ধির কথা, খালি অংকের ভাষায় লিখতে গিয়ে অমন খটমট রূপ হয়েছে। ধাপে ধাপে হজম করা যাক।

প্রথমে অংক-টংক ভুলে এই সহজ প্রশ্নটার উত্তর দাও। একটা জালি কয়েন আছে তার একদিকের probability হল 0.1 (অতএব অন্যদিকেরটা 0.9)। কিন্তু তুমি জানোনা ভারি দিকটায় কী আছে, head নাকি tail. সুতরাং আমাদের hypothesis দুটো হল

$$H_0 : P(\text{head}) = 0.1 \text{ আর } H_1 : P(\text{head}) = 0.9.$$

এই কয়েনটাকে এবার 10 বার টস করে এই data পাওয়া গেল--

H,H,H,T,H,H,T,H,H,H.

তবে সহজ বুদ্ধিতে বল তো কোন hypothesis-টা ঠিক বলে মনে হচ্ছে? এত বেশীবার head পড়েছে দেখে স্বাভাবিকভাবেই H_1-এর স্বপক্ষে রায় দেব আমরা। এখানে যে যুক্তিটা আমাদের মনের মধ্যে কাজ করছে সেটা ঠিক MLE বার করার মত--

- একবার $P(\text{head}) = 0.1$ বসিয়ে দেখছি likelihood কত হয়,
- তারপর $P(\text{head}) = 0.9$ বসিয়ে দেখছি কত হয়।

এই দুটো likelihood-এর তুলনার ভিত্তিতেই আমরা রায় দিচ্ছি। এই তুলনাটা করা যায় ভাগ করে--

$$LR = \frac{\text{likelihood under } H_0}{\text{likelihood under } H_1}.$$

এখানে LR মানে হল likelihood ratio. যদি এই LR-টা বেশ বড় হয় তবে আমরা H_0-র পক্ষে রায় দেব, ছোটো হলে H_1-এর স্বপক্ষে। এটুকু বুঝতে কোনো অসুবিধা নেই। কিন্তু ধরো যদি head আর tail-এর সংখ্যা সমান সমান হত? তবে দুটো likelihood-ই সমান হত (মানে $LR = 1$)। সেক্ষেত্রে তুমি দ্বিধাগ্রস্ত হয়ে পড়তে। এবার মনে রেখো যে type I error হল type II error-এর চেয়ে বেশী বিপজ্জনক। যেমন ধরো হয়তো বলা আছে যে যদি H_1 ঠিক হয়, আর তুমি ভুল করে H_0 বলে থাকো, তবে তোমার দশহাজার টাকা জরিমানা হবে। কিন্তু যদি উল্টো ভুলটা করো, তবে তৎক্ষণাৎ গর্দান। তাহলে দেখবে তোমার দ্বিধা চমৎকার কেটে গেছে, এবং তুমি H_0-এর স্বপক্ষে রায় দিচ্ছ। শুধু তাই নয় যদি দ্যাখো LR-টা 1-এর থেকে সামান্য কম, তাহলেও তুমি সাহস করে H_0-কে reject করতে পারবে না, কারণ দশহাজার টাকার মায়া যদি বা ছাড়া যায়, প্রাণের মায়া ছাড়া তার চেয়ে অনেক কঠিন। সুতরাং ভরসা করে H_0-কে reject তুমি তখনই খালি করবে যখন LR-টা খুবই ছোটো হবে। কতটা ছোটো সেটাকেই $LR < c$ হিসেবে লেখা হয়েছে। অনুভব করতে পারছ যে এটা করাই সবচেয়ে বুদ্ধিমানের কাজ। সেই কথাটাই অংকের ভাষায় লিখেছে যে এই test-টার power হল maximum.

Example 23: State Neyman-Pearson theorem to obtain the best critical region for simple hypothesis. [2] **(2008.5a)**
SOLUTION: আগের অংকটাই। ■

Exercise 6: State Neyman-Pearson's theorem.[2] **(2012.5bp1)**
HINT:
একই অংক আরেকবার। ■

37.2 হাতেকলমে

এবার আমরা Neyman-Pearson theorem-এর প্রয়োগ দেখব। প্রতি ক্ষেত্রে কিছু data দেওয়া থাকবে $X_1, ..., X_n$, যারা iid. একটা H_0 আর একটা H_1 দিয়ে দেবে, দুটোই simple. কোনো level of significance অবশ্য উল্লেখ করে দেবে না, সেটা আমাদেরই কিছু একটা ধরে নিতে হবে, যেমন α. তার জন্য Neyman-Pearson theorem লাগিয়ে MP test বার করতে হবে।

37.2.1 Normal mean

Example 24: Apply Neyman-Pearson's theorem to construct a test of the null hypothesis $H_0 : m = m_0$ against the alternative $H_1 : m = m_1$ for a Normal(m, σ) population where σ is known, and m_0 and m_1 are two given unequal numbers.[6] **(2005.9b)**
SOLUTION: এই প্রশ্নে বলতে ভুলে গেছে যে, data-টা কী। সেটা আমরা ধরে নেব--

> Let our data be $X_1, ..., X_n$ iid $N(m, \sigma)$, where $m \in \{m_0, m_1\}$ is unknown and $\sigma > 0$ is known.

এবার likelihood function-টা লিখে নেব। যেহেতু X_i-গুলোর pdf আছে, আর আমাদের data হল iid, তাই pdf-গুলোর গুণফলই হল likelihood function–

> Likelihood function is
>
> $$L(m) = \prod_1^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(X_i - m)^2}{2\sigma^2}\right) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum_1^n (X_i - m)^2}{2\sigma^2}\right).$$
>
> Consider a test of level α which rejects H_0 iff
>
> $$\frac{L(m_0)}{L(m_1)} < c$$
>
> for some c.
>
> NP theorem says that this test is MP for level α.

এইবার জিনিসটাকে সহজ করে লিখব--

> Now
>
> $$\begin{aligned} & \frac{L(m_0)}{L(m_1)} < c \\ \iff \quad & \frac{\exp\left(-\frac{\sum_1^n (X_i - m_0)^2}{2\sigma^2}\right)}{\exp\left(-\frac{\sum_1^n (X_i - m_1)^2}{2\sigma^2}\right)} < c \\ \iff \quad & \exp\left[-\tfrac{1}{2\sigma^2}\left(\sum_1^n (X_i - m_0)^2 - \sum_1^n (X_i - m_1)^2\right)\right] < c \\ \iff \quad & \sum_1^n (X_i - m_1)^2 - \sum_1^n (X_i - m_0)^2 > \underbrace{-2\sigma^2 \log c}_{c',\ \text{say}} \quad \left[\because e^x \text{ is increasing function}\right] \end{aligned}$$
>
> লক্ষ কর এখানে মাইনাস থাকায় '$<$'-টা কেমন '$>$' হয়ে গেল
>
> $$\begin{aligned} \iff \quad & \sum_1^n \left[(X_i^2 - 2m_1 X_i + m_1^2) - (X_i^2 - 2m_0 X_i + m_0^2)\right] > c' \\ \iff \quad & 2(m_0 - m_1) \sum_1^n X_i + (m_1^2 - m_0^2) > c' \\ \iff \quad & 2(m_0 - m_1) \sum_1^n X_i < \underbrace{c' - (m_1^2 - m_0^2)}_{c'',\ \text{say}} \end{aligned}$$

এবার $2(m_0 - m_1)$ দিয়ে ভাগ করতে চাই, যাতে বাঁ দিকে খালি $\sum X_i$ থাকে। ভাগটা করতে অসুবিধা নেই, কারণ বলাই আছে যে, $m_0 \neq m_1$. কিন্তু এই ভাগের ফলে inequality-টার পরিবর্তন হবে কিনা সেটা নির্ভর করবে $m_0 - m_1$-টা > 0 নাকি < 0, তার উপরে। সেই অনুযায়ী দুটো কেস হবে। দুটো কেসের যুক্তিই খুবই একইরকম, একটা ভালো করে বুঝলে অন্যটা নিজেই করতে পারবে।

> Case I: If $m_0 < m_1$, then $m_0 - m_1 < 0$ and so our MP test rejects H_0 when
> $$\sum_1^n X_i > \frac{c''}{2(m_0 - m_1)} = c''', \text{ say.}$$

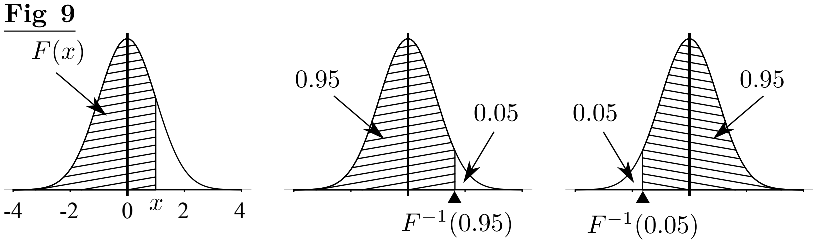
এইবার কাজ হল এই c'''-টাকে বার করা। মনে রেখো যে আমাদের test-টা র level আমরা নিয়েছিলাম α. আমরা চাই c'''-কে α দিয়ে প্রকাশ করতে।

> $\because$ The level of the test is α,
>
> $\therefore$ Under H_0,
> $$P\left(\sum_1^n X_i > c'''\right) = \alpha.$$

এইটা বুঝলে তো? আমরা চাই $P(\text{type I error}) = \alpha$. আর $P(\text{type I error})$ মানে হল H_0 ঠিক হলে H_0-কে ভুল করে reject করার probability. আমরা H_0-কে reject করছি কখন? যখন, $\sum_1^n X_i > c'''$.

> Under H_0, we know that $\sum_1^n X_i \sim N(nm_0, \sqrt{n}\sigma)$.
>
> $\therefore Z \equiv \frac{\sum_1^n x_i - nm_0}{\sqrt{n}\sigma} \sim N(0,1)$.
>
> Now
> $$\begin{aligned} & P\left(\sum_1^n X_i > c'''\right) = \alpha \\ \iff \quad & P\left(Z > \tfrac{c - nm_0}{\sqrt{n}\sigma}\right) = \alpha \\ \iff \quad & P\left(Z \leq \tfrac{c - nm_0}{\sqrt{n}\sigma}\right) = 1 - \alpha \\ \iff \quad & F\left(\tfrac{c - nm_0}{\sqrt{n}\sigma}\right) = 1 - \alpha, \end{aligned}$$
>
> where $F(\cdot)$ is the cdf of $N(0,1)$ distribution.

Fig 9



> Hence our test is to reject H_0 if and only if
>
> $$\sum_1^n X_i > nm_0 + \sqrt{n}\sigma F^{-1}(1-\alpha).$$

Fig 9 দ্যাখো। এই ছবিটা $\alpha = 0.05$ ধরে এঁকেছি, তাই $1-\alpha = 0.95$.
অন্য কেসটাও একদম একইরকম, খালি inequality-টা উল্টে যাবে। তাই "similarly" লিখে ছেড়ে দিতে পারি--

> Case 2: If $m_0 > m_1$, then similarly our test rejects H_0 if and only if
>
> $$\sum_1^n x_i < nm_0 - \sqrt{n}\sigma F^{-1}(1-\alpha).$$

■

37.2.2 Normal variance

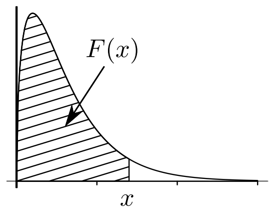
Example 25: For the normal(m,s) population where m is known, test the null hypothesis $H_0 : s = s_0$ against the alternative $H_1 : s = s_1$ on the basis of a sample $(x_1, x_2, ..., x_n)$ from the population.[7] **(2004.9b)**
SOLUTION: এখানে বলে দিয়েছে data-টা কী। সেটা লিখেই আমরা শুরু করব--

> Let our data be $X_1, ..., X_n$ iid $N(m,s)$, where m is known and $s \in \{s_0, s_1\}$ is unknown.

এবার likelihood function লেখার পালা। সেটা ঠিক আগের অংকের মতই--

> Likelihood function is
>
> $$L(s) = \prod_1^n \frac{1}{s\sqrt{2\pi}} \exp\left(-\frac{(x_i-m)^2}{2s^2}\right) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum_1^n (x_i-m)^2}{2s^2}\right).$$
>
> Consider a test of level α which rejects H_0 iff
>
> $$\frac{L(s_0)}{L(s_1)} < c$$
>
> for some c.
> NP theorem says that this test is MP for level α.

এবার সহজ করে লেখার চেষ্টা করব--



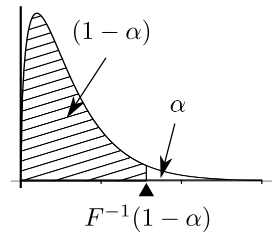


Fig 10

Now

$$\begin{aligned} & \frac{L(s_1)}{L(s_0)} > c \\ \iff & \cdots \end{aligned}$$

এই জায়গাটা তোমার করার জন্য রেখে দিলাম।

$$\iff \sum_1^n (x_i - m)^2 \left(\frac{1}{s_0^2} - \frac{1}{s_1^2}\right) > c''.$$

যেহেতু বলে দেয় নি যে, s_0 আর s_1-এর মধ্যে কে বড়, তাই দুটো কেসে ভেঙে এগোতে হবে--

Case 1: If $s_0 < s_1$, then $\frac{1}{s_0^2} - \frac{1}{s_1^2} > 0$ and so our MP test rejects H_0 iff

$$\underbrace{\sum_1^n (x_i - m)^2}_{T, \text{ say}} > \underbrace{\left(\tfrac{1}{s_0^2} - \tfrac{1}{s_1^2}\right)^{-1} c''}_{c''', \text{ say}}.$$

এবার c'''-কে α দিয়ে প্রকাশ করতে হবে। তার জন্য এই তথ্যটা কাজে লাগবে--

Under H_0, $\frac{T}{s_0^2} \sim \chi^2_{(n)}$.

$\because$ The level of the test is α,

$\therefore$ Under H_0,

$$P(T > c''') = \alpha \iff P\left(\tfrac{T}{s_0^2} > \tfrac{c'''}{s_0^2}\right) = \alpha \iff P\left(\tfrac{T}{s_0^2} \leq \tfrac{c'''}{s_0^2}\right) = 1-\alpha \iff F\left(\tfrac{c'''}{s_0^2}\right) = 1-\alpha,$$

where $F(\cdot)$ is the cdf of $\chi^2_{(n-1)}$ distribution.

So our test is to reject H_0 when

$$\sum_1^n (x_i - m)^2 > s_0^2 F^{-1}(1-\alpha).$$

Fig 10 দ্যাখো। অন্য কেসটা একইরকম।

Case 2: If $s_0 > s_1$, then similarly our test is to reject H_0 when

$$\sum_1^n (x_i - m)^2 < s_0^2 F^{-1}(\alpha).$$

"Similarly" বলে ছেড়ে দিয়েছি বটে, কিন্তু তুমি যেন রাফে কষে মিলিয়ে নিও। বিশেষতঃ F^{-1}-এর মধ্যে $1-\alpha$-র জায়গায় কী করে α এল সেটা বুঝে নিও। ■

Example 26: Using Neyman-Pearson's theorem construct a test of null hypothesis $H_0 : \sigma = \sigma_0$ against the hypothesis $H : \sigma = \sigma_1$ for the normal(m, σ) population.[5] **(2012.5bp2)**

SOLUTION: এই অংকে একটা গুরুতর গাফিলতি আছে। বলে দেয় নি m-টা জানা নাকি অজানা। কিছু বলে দেওয়া না থাকলে অজানা ধরে নেওয়াটাই স্বাভাবিক, কিন্তু এখানে তা ধরলে চলবে না, কারণ $m \in \mathbb{R}$ যদি অজানা হয় তবে H_0 আর H_1 দুজনেই composite hypothesis. সেক্ষেত্রে Neyman-Pearson theorem সরাসরি প্রযোজ্য নয়। যেহেতু ওই theorem-টাই লাগাতে বলেছে, তাই এখানে ধরে নিতে হবে যে, m জানা আছে। এবার এটা আগের অংকটাতেই পরিণত হল। ■

37.2.3 Composite hypothesis

Neyman-Pearson theorem যে composite hypothesis-এর বেলায় সরাসরি প্রযোজ্য নয় সে তো theorem-টার শর্ত থেকেই বোঝা আছে, কারণ ওখানে বলাই আছে যে H_0 আর H_1 দুজনকেই simple hypothesis হতে হবে। শুধু যে theorem-এর শর্ত নিয়েই সমস্যা তাই নয়, আসলে H_1 যদি composite হয়, তবে most powerful-এর ধারণাটাই ধোপে টেঁকে না। কেন শোনো। Power মানে হল H_1 ঠিক হলে পরে H_0-কে reject করার probability. যেহেতু H_1 হল composite, তাই Θ_1-এর মধ্যে অন্ততঃ দুটো element আছে, ধরো θ_1 আর θ_2. H_1 ঠিক হলে পরে θ-র value কিন্তু এখনও অজানা, যেমন $\theta = \theta_1$-ও হতে পারে আবার $\theta = \theta_2$-ও হতে পারে। এবং এই দুইক্ষেত্রে দুরকম power আসবে। তুমি কোনটাকে maximum করার চেষ্টা করবে? θ_1-এ যে test-টার power-টা maximum হল, θ_2-তে হয়তো সে আরেকটা test-এর কাছে মার খেয়ে গেল।

কিন্তু খুব বিশেষ কিছু ক্ষেত্রে এমন একটা জব্বর test বার করা যায় যেখানে Θ_1-এর মধ্যে প্রতিটা value-তেই power-টা maximum হয়। এরকম জব্বর test খুব কম সময়েই পাওয়া যায়। যখন পাওয়া যায় তখন এদের বলে **uniformly most powerful test** বা সংক্ষেপে **UMP test**. এরকম একটা বিশেষক্ষেত্রের আলোচনা রয়েছে নীচের অংকে। এরকম test যদি থাকে, তবে Neyman-Pearson theorem দিয়েই এদের বার করা যায়। সেটাই এবার শিখব।

Example 27: For a normal(m, σ) population with known σ, construct by means of Neyman-Pearson theorem, a test for null hypothesis

$$H_0 : m = m_0 \text{ against the alternative } H_1 : m > m_0.$$

[7] **(2009.5a)**

SOLUTION: এখানে H_1-টা composite কারণ $\Theta_1 = (m_0, \infty)$, যেটা মোটেই singleton set নয়। তাই সরাসরি Neyman-Pearson theorem তো লাগাতে পারব না। সেই কারণে একটা কৌশল করব। Θ_1-এর থেকে খালি যে কোনো একটা প্রতিনিধি তুলে নেব, যার নাম দিলাম ধরো m_1.

First we take any $m_1 > m_0$ and consider testing $H_0 : m = m_0$ against $H_1' : m = m_1$.

এখন দুটো hypothesis-ই simple, তাই Neyman-Pearson theorem লাগানো যাবে। এবার আমাদের পরিকল্পনাটা বলি। আমরা Neyman-Pearson theorem লাগিয়ে একটা MP test পাব। যদি দেখা যায় যে, সেই test-টার মধ্যে কোথাও m_1 নেই (মানে যেখানে যা m_1 ছিল সব যদি কাটাকাটি হয়ে যায়), তবে বুঝতে হবে সেই test-টা আসলে যেকোনো $m_1 \in \Theta_1$-এর জন্যই MP, অর্থাৎ কিনা UMP. অবশ্য যদি m_1-টা কাটাকাটি না হয়ে যায়, তবে কোনো test-ই power-এর যুদ্ধে অবিসংবাদিতভাবে জিততে পারছে না। সেক্ষেত্রে এটা প্রমাণ হয়ে যাচ্ছে যে, কোনো UMP test-এর অস্তিত্ব নেই। এই অংকে অবশ্য দিব্যি কাটাকাটি হয়ে যাবে। সেটা দেখানোই এখানে আমাদের প্রধান কাজ।
এর জন্য আমাদের ঠিক 327 নম্বর পাতার Example 24-র ধাপগুলোই করে যেতে হবে। সেই ধাপগুলো আর নতুন করে লিখছি না, সেগুলো তুমি লিখে নিও। খালি একটাই পার্থক্য--ওখানে দুটো কেস ছিল, কিন্তু এখানে জানাই আছে যে $m_0 < m_1$, তাই খালি প্রথম কেসটাই লাগবে। শেষে পাবে--

> Hence our test is to reject H_0 if and only if
>
> $$\sum_1^n x_i > nm_0 + \sqrt{n}\sigma F^{-1}(1-\alpha).$$

লক্ষ কর যে, এর মধ্যে m_1 কোথাও নেই। সুতরাং আমাদের কৌশলটা কাজ করেছে।

> Since this MP test does not depend on the value of $m_1 > m_0$, so the same test must be UMP against the alternative $H_1 : m > m_0$.

■

DAY 38
Likelihood Ratio Tests (exact)

গতকাল আমরা দেখলাম যে H_0 আর H_1 দুজনেই simple hypothesis হলে কীভাবে Neyman-Pearson theorem লাগিয়ে একটা ভালো test বার করা যায়। কিন্তু composite hypothesis-এর বেলায় সমস্যা বাঁধে। কোনো কোনো ক্ষেত্রে (যেখানে UMP test থাকে) Neyman-Pearson theorem লাগানো যায় বটে, কিন্তু সেরকম ক্ষেত্র কমই পাওয়া যায়। অথচ দ্যাখো, Neyman-Pearson theorem-এর মূল ধারণাটা কিন্তু বেশ ভালোই ছিল--H_0 ঠিক হলে likelihood কত হয়, আর H_1 ঠিক হলেই বা কত হয়, এ দুয়ের তুলনা। সেই একই কায়দা আমরা কিন্তু composite hypothesis-এর ক্ষেত্রেও লাগাতে পারি, তাকে বলে **likelihood ratio test** বা **LRT.** দুঃখের ব্যাপার এই যে, এখন আর Neyman-Pearson theorem-এর মত কোনো theorem নেই যেটা গ্যারান্টি দিয়ে বলবে এই test-টা সবচেয়ে ভালো। কিন্তু সবচেয়ে ভালো নাও যদি হয়, বহুক্ষেত্রেই এই কায়দায় বেশ ভালো test বানানো যায়। তাই statistics-এ likelihood ratio test একটা গুরুত্বপূর্ণ স্থান অধিকার করে আছে।

38.1 জিনিসটা কী?

একটা কলেজে ছেলেদের আর মেয়েদের মধ্যে পড়াশোনা নিয়ে বেজায় রেষারেষি। কোনো পরীক্ষা হলেই ছেলেদের থেকে ফার্স্ট হয় নাকি মেয়েদের থেকে হয়, তা নিয়ে প্রচুর জল্পনা চলতে থাকে। বুঝতেই পারছ যে রেষারেষিটা কোনো একটা ছেলে আর একটা মেয়ের মধ্যে তুলনার প্রশ্ন নয়। যদি তা হত, তাহলে দুজনের নম্বরের তুলনা করলেই সে প্রশ্নের নিরসন করা যেত। কিন্তু এখানে লড়াই হচ্ছে ছেলেদের দলের সঙ্গে মেয়েদের দলের। অবশ্যই এমন মেয়ে পাওয়া যাবে যে কোনো একটা ছেলের চেয়ে কম নম্বর পেয়েছে। কিন্তু তা বলে মেয়েরা হার মানতে রাজী হবে না। তারা হই হই করে উঠবে এই বলে যে, এমন ছেলেও আছে যে কোনো একটা মেয়ের থেকে কম নম্বর পেয়েছে। এসব ক্ষেত্রে তুলনাটা করা দরকার মেয়েদের মধ্যে সবচেয়ে বেশী নম্বরের সঙ্গে ছেলেদের সবচেয়ে বেশী নম্বরের। একইরকম ব্যাপার আমরা খেলার জগতেও আকছার দেখে থাকি।

ভারত-পাকিস্তানের টেস্ট ম্যাচে যখন আমরা বলি ভারত পাকিস্তানকে হারিয়েছে, তার মানে ভারতের সেরা ক্রিকেটারদের নিয়ে তৈরী টীম পাকিস্তানের সেরাদের নিয়ে তৈরী টীমকে হারিয়েছে।
ঠিক এই যুক্তিটাই আমরা এখানে খাটাব, hypothesis testing-এর বেলায়। ধরো আমরা test করছি

$$H_0 : \theta \in \Theta_0 \text{ বনাম } H_0 : \theta \in \Theta_1.$$

আমরা গত অধ্যায়ে Neyman-Pearson theorem শিখেছিলাম। সেটা দিয়ে simple H_0-এর against-এ simple H_1-এর test করা যেত। কায়দাটা ছিল সহজ--

> H_0 হলে likelihood কত হয় দ্যাখো, আবার H_1 হলেই বা কত হয় দ্যাখো। যদি H_0-এর বেলায় অনেকটাই ছোটো হয় তবে H_1-এর পক্ষে রায় দাও, নইলে H_0-এর পক্ষে। ব্যস্!

এখানে likelihood-টা যেন পরীক্ষার নম্বর, আর Θ_0 হল ছেলেদের দল, এবং Θ_1 হল মেয়েদের দল। যেহেতু দুটোই simple hypothesis, তার মানে ছেলেদের দলে এখানে খালি একটাই ছেলে, আর মেয়েদের দলেও খালি একটাই মেয়ে। মানে একটা ছেলের সঙ্গে একটা মেয়ের তুলনা হচ্ছে। তাই সরাসরি তাদের নম্বর তুলনা করলেই চলছিল। সেটাই ছিল Neyman-Pearson theorem-এর মূল কথা। কিন্তু যদি H_0, H_1-এর মধ্যে একজনও composite হয়, তবেই সেই দলাদলির সমস্যা। সেক্ষেত্রে আমরা এক্ষুণি যে যুক্তির কথা বললাম সেই পথে এগোব। H_0 ঠিক হলে likelihood-টা সবচেয়ে বেশী কত হতে পারে দেখব। সেইটা হল H_0 দলের সর্বশ্রেষ্ঠ প্রতিনিধি। একইভাবে H_1 ঠিক হলেই বা likelihood-টা সবচেয়ে কত উঠতে পারে দেখব। এবার সেই দুই প্রতিনিধির মধ্যে তুলনা করব। "সবচেয়ে বেশী" বললে সাধারণভাবে maximum-এর কথাই মনে আসে। কিন্তু তার চেয়ে বেশী general হল supremum (যেমন $(0, 1)$-র কোনো maximum নেই, কিন্তু supremum আছে, 1). সুতরাং H_0-এর প্রতিনিধি আমরা নেব

$$\sup_{\theta\in\Theta_0} L(\theta)$$

আর H_1-এর বেলায়

$$\sup_{\theta\in\Theta_1} L(\theta).$$

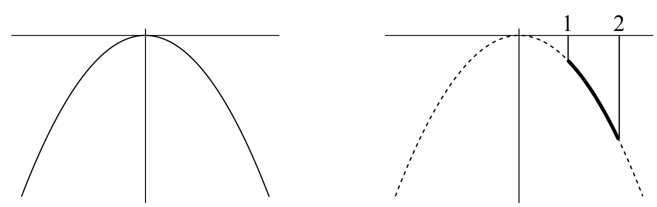
সুতরাং এবার আমরা test-টা করব এইভাবে--

> যদি $\sup_{\theta\in\Theta_0} L(\theta)$ দেখা যায় যে $\sup_{\theta\in\Theta_1} L(\theta)$-এর চেয়ে অনেকটাই ছোটো, তবে H_1-র স্বপক্ষে রায় দেব, নইলে H_0-এর স্বপক্ষে।

যেহেতু likelihood-দের (supremum-এর) ratio নিয়ে কাজ হচ্ছে, তাই এই রকম test-কে আমরা বলি likelihood ratio test বা সংক্ষেপে LRT.
শুনতে ভালোই লাগছে, তবে ওই supremum দুটো বার করতে একটু বাড়তি খাটনি হবে। সেই খাটনি খানিকটা কমানোর একটা পথ আছে। সেটা আবার ছেলে-মেয়েদের উদাহরণ দিয়ে বুঝে নিই। ছেলেদের সর্বোচ্চ নম্বর আর মেয়েদের সর্বোচ্চ নম্বর দুটোই বার না করে, যে কোনো একটা বার কর। তার সাথে তুলনা কর পুরো ক্লাসের সর্বোচ্চ নম্বরের। মানে অংকের ভাষায়

> $\sup_{\theta\in\Theta_0} L(\theta)$-এর সঙ্গে $\sup_{\theta\in\Theta} L(\theta)$-এর।

Fig 11



হঠাৎ দেখে মনে হতে পারে যে, এতে লাভটা কী হল। সেই তো দুটো supremum-ই বার করতে হচ্ছে! ঠিক কথা, কিন্তু অনেক সময়েই $\sup_{\theta\in\Theta} L(\theta)$ বার করা $\sup_{\theta\in\Theta_1} L(\theta)$ বার করার থেকে সহজতর হয়। কারণ Θ হল θ-র যাবতীয় value-র set, সাধারণতঃ সেটা $\mathbb{R}$ বা সেরকম ভদ্রকিছু হয়। অপরপক্ষে Θ_1 হল তার একটা subset, যেটা ততটা ভদ্র নাও হতে পারে। নীচের উদাহরণটা দেখলেই বুঝবে একটা subset-এর উপরে supremum বার করা কেন বেশী কঠিন হতে পারে।

Example 28: ধরো $f(x) = -x^2$. পুরো $\mathbb{R}$-এর উপরে এর maximum কত? যদি x খালি $[1,2]$-র মধ্যে থাকতে পারে, তবে?

SOLUTION: Fig 11 দ্যাখো। প্রথম ক্ষেত্রে খালি differentiate করে 0-র সমান করলেই পাবে $f'(x) = -2x = 0$, মানে $x = 0$. এবার second derivative-টা পরীক্ষা করে দ্যাখো, $f''(0) = -2 < 0$. সুতরাং উত্তর হল $x = 0$-তে maximum হচ্ছে। কিন্তু যখন $x \in [1,2]$ তখন এই সহজ কায়দায় চলবে না। তখন দুই প্রান্ত $x = 1$ আর $x = 2$-তেও $f(x)$-এর value পরীক্ষা করে দেখতে হবে। ■

এই কারণে LRT-র সংজ্ঞা বলা হয়ে থাকে এইভাবে--

DEFINITION: **Likelihood ratio test (LRT)**

Suppose that we have a data set $X_1, ..., X_n$ with likelihood $L(\theta)$ where $\theta \in \Theta$. We want to test

$$H_0 : \theta \in \Theta_0 \text{ against } H_1 : \theta \in \Theta_1,$$

where $\Theta = \Theta_0 \cup \Theta_1$ and $\Theta_0 \cap \Theta_1 = \phi$.
Then a likelihood ratio test rejects H_1 if and only if

$$\frac{\sup_{\theta\in\Theta_0} L(\theta)}{\sup_{\theta\in\Theta} L(\theta)} < c,$$

where c is a constant determined by the level of the test.

পুরো জিনিসটা বেশ গালভরা কিন্তু আগের গল্পটা বুঝে থাকলে অসুবিধা হওয়া উচিত নয়। একেবারে শেষে যে লিখেছি c-কে level ব্যবহার করে বার করা যাবে, সেটা ঠিক গত অধ্যায়ে যেভাবে Neyman-Pearson theorem-এর জন্য করছিলাম সেরকমই। নীচের অংকগুলো যখন কষব, তখন স্পষ্ট হবে।
আমরা যেসব অংক দেখব সেখানে ব্যাপারটা আরও সহজ হয়ে যাবে, supremum দুটো নিয়ে কোনোই কষ্ট করতে হবে না। সেটা কী করে হবে বোঝার জন্য likelihood ratio-টার নীচেরতলার supremum-টার দিকে তাকাও, মানে $\sup_{\theta\in\Theta} L(\theta)$. এখানে পুরো Θ-র উপরে কাজ হচ্ছে। পুরো Θ-র উপরে θ-র যে value-র জন্য $L(\theta)$ সবচেয়ে বেশী হয়, তাকে যে MLE বলে সে তো জানোই। সুতরাং MLE-টাকে যদি $\hat{\theta}$ বলি, তবে নীচের তলার supremum-টা আসলে $L(\hat{\theta})$ ছাড়া আর কিছুই নয়। অতএব MLE-র ফর্মুলাটা যদি তোমার জানা থাকে, তবে নীচের তলার supremum-টা অনায়াসে বেরিয়ে যাবে।
উপরতলার supremum-টা নিয়ে ঝামেলা হতে পারে, তবে যদি H_0 হয় simple, মানে Θ_0 একটা singleton set হয়, ধরো $\Theta_0 = \{\theta_0\}$, likelihood-টা খালি একটাই value নেবে, $L(\theta_0)$. সুতরাং তার supremum-ও হবে সেই $L(\theta_0)$-ই। অবশ্য আমরা এমন অংকও করব যেখানে H_0-টা মোটেই simple নয়। কিন্তু তাও দেখবে উপরতলার supremum-টা অনায়াসেই বার করা যাচ্ছে। সেটা যথাসময়ে অংকগুলো করার সময়েই দেখবে। আগে একটা উদাহরণ দেখে নিই যেখানে H_0-টা simple.

38.2 Normal mean

Example 29: Use the method of likelihood ratio testing to explain how would you test the

hypothesis for the mean of normal(m,σ) population, where σ is known, against no specific alternatives.[7]
(2011.5c)

SOLUTION: এই অংকটায় বর্ণনাটা অসম্পূর্ণ। এটুকু বোঝা যাচ্ছে যে একটা test করতে বলা হচ্ছে, এবং সে কাজে likelihood ratio test লাগাতে হবে সেটাও বলা আছে। কিন্তু এটা বোঝা যাচ্ছে না যে, ঠিক কী hypothesis-কে test করতে হবে, মানে H_0-ই বা কী, আর H_1-ই বা কী, আর আমাদের data-ই বা কী। এক জায়গায় normal(m,σ) বলা আছে, এবং বলেছে σ is known. কিন্তু "the hypothesis for the mean" বলার মানে কী? এখানে "the hypothesis" শুনে মনে হয় যেন একটাই খালি hypothesis সম্ভব। তা মোটেই নয়। আসলে এখানে বলতে চাইছে সেই hypothesis-টার কথা যেটার জন্য LRT বার করাটা সবচেয়ে সহজ।
যাই হোক যেহেতু আমরা প্রশ্নকর্তা নই, তাই প্রশ্নের শব্দগুলো হুবহুই তুলে দিয়েছি। প্রশ্নের গাফিলতি আমরা উত্তরটা লেখার সময়ে পুষিয়ে নেব। প্রথমেই লিখে নেব আমরা সত্যিই কী করতে চলেছি। সবার শুরুতে data-র বর্ণনা--

> Step 0: Let $X_1,...,X_n$ be iid $N(m,\sigma)$,

তারপর parameter এবং parameter space-এর বিবরণ--

> where $m \in \mathbb{R}$ is unknown and $\sigma > 0$ is known.

তারপর লিখি H_0 কী, আর H_1-ই বা কী।

> We want to test $H_0 : m = m_0$ against $H_1 : m \neq m_0$, where $m_0 \in \mathbb{R}$ is any given number.

এইটা কী করে পেলাম? এখানে বলেছিল mean-টাকে test করতে, মানে H_0-কে হতে হবে m-এর বিষয়ে কিছু একটা বক্তব্য। এখানে বক্তব্য নানারকম হতে পারে, যেমন $m \in (0,1)$ বা $m \in \mathbb{Q}$ ইত্যাদি। সেগুলোর জন্যও LRT বার করা যেতে পারে, কিন্তু বেজায় শক্ত। সবচেয়ে সহজ হল যেখানে H_0 হবে simple, মানে "$m =$ কোনো জানা সংখ্যা"-জাতীয়। কারণ এইক্ষেত্রে LRT-র উপরতলার supremum-টা বেরিয়ে যাবে অনায়াসে। তা, আমরা সেই "কোনো জানা সংখ্যা"-র জায়গায় m_0 নিয়েছি। তাহলে H_1 হবে $m \neq m_0$. এখানে বলেছে "no specific alternative". এ কথাটার কোনো মানে হয় না। অনুমান করা যায় যে, প্রশ্নকর্তা বলতে চেয়েছিলেন "no specific direction", মানে খালি $m > m_0$ বা খালি $m < m_0$ নয়। তুমি যদি জীবনে কখনও প্রশ্নকর্তার ভূমিকায় অবতীর্ণ হও, তখন যেন এরকম অসম্পূর্ণ উদ্ভট ভাষা ব্যবহার কোরো না।
এবার তবে আসল কাজ শুরু করি। প্রথমেই লিখে নেব likelihood function-টা--

> The likelihood function is
>
> $$\begin{aligned} L(m) &= \prod_1^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(X_i - m)^2\right) \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_1^n (X_i - m)^2\right) \end{aligned}$$

আমাদের এবারকার কাজ হল $\sup_{\theta\in\Theta_0} L(\theta)$ বার করা। এখানে θ-র ভূমিকায় আছে m, আর $\Theta_0 = \{m_0\}$. যেহেতু Θ_0-টা একটা singleton set, সুতরাং supremum বার করার কিছু নেই--

Step 1: Here $\Theta_0 = \{m_0\}$ and $\Theta = \mathbb{R}$. So

$$\sup\{L(m) \ : \ m \in \Theta_0\} = \sup\{L(m_0)\} = L(m_0).$$

সুতরাং উপরতলার supremum-টার হিল্লে হয়ে গেল। এবার নীচের তলা। এখানে MLE দিয়ে কাজ সেরে ফেলা যাবে, কারণ--

Step 2: $\sup\{L(m) \ : \ m \in \Theta\} = \sup\{L(m) \ : \ m \in \mathbb{R}\} = L(\widehat{m})$, where $\widehat{m}$ is the MLE of m.

From standard result we know that $\widehat{m} = \bar{X}$.

এই MLE-র ফর্মুলাটা মনে রাখা ভালো। এখানে সেটা নতুন করে বার করতে হলে উত্তরটা বড্ড লম্বা হয়ে যেত।

So $\sup\{L(m) \ : \ m \in \Theta\} = L(\bar{X})$.

ব্যস্, নীচের তলার supremum-ও বেরিয়ে গেল। এইবার LRT-টা বানানো যাক। এর জন্য কাজ করতে হবে $L(m_0)/L(\bar{X})$ নিয়ে। মনে রেখো যে এটা খুব ছোটো হলে আমরা H_0-কে reject করব। এটাকে সাজিয়ে গুছিয়ে অনেক সহজ করে তোলা যায়--

Step 3: So the LRT rejects H_0 iff

$$\frac{L(m_0)}{L(\bar{X})} < c$$

$$\iff \frac{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - m_0)^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - \bar{X})^2\right)} < c$$

$$\iff \exp\left(-\frac{1}{2\sigma^2}\sum_1^n \left[(X_i - m_0)^2 - (X_i - \bar{X})^2\right]\right) < c$$

$$\iff \sum_1^n \left[(X_i - m_0)^2 - (X_i - \bar{X})^2\right] > \underbrace{\log c}_{c',\ \text{say}} \quad \left[\because e^x \text{ is an increasing function}\right]$$

...

এর পরের কয়েকটা মামূলী ধাপ আর লিখলাম না।

$$\iff n(\bar{X} - m_0)^2 > c''$$

$$\iff |\bar{X} - m_0| > \underbrace{\sqrt{c''/n}}_{c''',\ \text{say.}}$$

একটু চিন্তা করলেই বুঝবে যে এত অংক কষে আমরা যা পেলাম সেটা খুবই স্বাভাবিক। আমাদের কাজ ছিল $m = m_0$ কিনা দেখা। এখানে m আমাদের অজানা হলেও $\bar{X}$-টা m-এর কাছাকাছি কিছু একটা হবে। সুতরাং আমরা যেন $\bar{X}$-কে m-র প্রক্সি হিসেবে ব্যবহার করছি। তাই m-কে m_0-এর সঙ্গে তুলনা না করে $\bar{X}$-কে m_0-এর সঙ্গে তুলনা করছি।

We need to choose c''' such that $P(|\bar{X} - m_0| > c''') = \alpha$ under H_0, where α

is the level of significance.

এইবার c'''-টা কী করে বার করব সেটা লেখা যাক--

Step 4: Now, under H_0, we know that $\bar{X} \sim N\left(m_0, \frac{\sigma}{\sqrt{n}}\right)$.

$\therefore \frac{\sqrt{n}(\bar{X}-m_0)}{\sigma} \sim N(0,1)$.

Thus

$$\begin{aligned} \alpha &= P(|\bar{X}-m_0| > c''') \\ &= P\left(\frac{\sqrt{n}}{\sigma}|\bar{X}-m_0| > \frac{\sqrt{n}c'''}{\sigma}\right) \\ &= 2\left[1 - F\left(\frac{\sqrt{n}c'''}{\sigma}\right)\right] \quad [\text{ where } F(\cdot) \text{ is the } N(0,1) \text{ CDF}] \end{aligned}$$

এইটা কী করে হল, একটু চিন্তা করে বুঝে নাও। ছবি আঁকলে সুবিধা হবে।

So $c''' = \frac{\sigma}{\sqrt{n}} F^{-1}\left(1-\frac{\alpha}{2}\right)$.

■

এরপরের অংকটাও একইরকম, খালি এবার σ-টা জানা নেই। সেটাতে হাত দেওয়ার আগে ভালো করে আগের অংকের সমাধানের কাঠামোটা বুঝে নেওয়া যাক। আমরা পাঁচ ধাপে এগিয়েছিলাম, step 0 থেকে step 4. এই ধাপগুলো LRT-র যেকোনো অংকেই থাকবে।

- Step 0-তে বলে নেব set up-টা। এখানে আসবে
  1. data-র বর্ণনা,
  2. distribution-টা কী, তার parameter-গুলো কী কী, এবং parameter space-ই বা কী,
  3. H_0 আর H_1 কী,
  4. Likelihood function-টা কী।
- এর পর Step 1-এ বার করব উপরতলার supremum-টা।
- Step 2-তে বার করব নীচের তলার supremum-টা।
- Step 3-তে likelihood ratio-টাকে দলাইমলাই করে যথাসম্ভব সহজ রূপে আনার চেষ্টা করব। এই ধাপটা অনেকের একটু অসুবিধা হয়, কারণ "সহজ রূপ" বলতে কী বোঝানো হচ্ছে সেটা সহজে বোঝা যায় না। আমাদের উদ্দেশ্য থাকে "$LR < c$" শর্তটাকে $T < c'$ বা $T > c'$ বা $|T| < c'$ বা $|T| > c'$ জাতীয় কিছু একটাতে পরিণত করা, যেখানে T হল $X_1, ..., X_n$-এর এমন একটা function, যার null distribution আমাদের জানা আছে।
- Step 4-এ বার করব critical value-টা।

একই কাঠামো আমরা ব্যবহার করব নীচের অংকটাতেও।

Example 30: Apply the method of likelihood ratio testing to develop a method of testing hypothesis $H_0 : m = m_0$ for a normal(m, σ) population.[7] **(2006.5b)**

SOLUTION:

Step 0: Let $X_1, ..., X_n$ be iid $N(m, \sigma)$, where $m \in \mathbb{R}$ and $\sigma > 0$ are unknown. Thus here $\Theta = \{(m, \sigma) \ : \ m \in \mathbb{R}, \sigma > 0\}$.

We want to construct LRT for

$$H_0 : m = m_0 \text{ against } H_1 : m \neq m_0,$$

where $m_0 \in \mathbb{R}$ is a given number. So $\Theta_0 = \{(m_0, \sigma) \ : \ \sigma > 0\}$.

The likelihood function is

$$L(m, \sigma^2) = \prod_1^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(X_i - m)^2\right) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_1^n (X_i - m)^2\right).$$

Step 1: Under H_0 the value of m is m_0, and the MLE of σ^2 is $\widehat{\sigma}_0^2 = \frac{1}{n}\sum_1^n (X_i - m_0)^2$. So the maximum value of the likelihood is $L(m_0, \widehat{\sigma}_0^2)$.

Step 2: In general the MLE's are $\widehat{m} = \bar{X}$ and $\widehat{\sigma}^2 = \frac{1}{n}\sum_1^n (X_i - \bar{X})^2$. So the maximum value of $L(m, \sigma^2)$ is $L(\widehat{m}, \widehat{\sigma}^2)$.

Step 3: So the LRT rejects H_0 iff

$$
\begin{aligned}
& \frac{L(m_0, \widehat{\sigma}_0^2)}{L(\widehat{m}, \widehat{\sigma}^2)} < c \\
\iff \quad & \left(\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2}\right)^{-n/2} < c \\
\iff \quad & \frac{\widehat{\sigma}^2}{\widehat{\sigma}_0^2} > \underbrace{c^{-2/n}}_{c', \text{ say}} \\
\iff \quad & \frac{\sum_1^n (X_i - m_0)^2}{\sum_1^n (X_i - \bar{X})^2} > c' \\
\iff \quad & \frac{\sum_1^n (X_i - \bar{X} + \bar{X} - m_0)^2}{\sum_1^n (X_i - \bar{X})^2} > c' \\
\iff \quad & \frac{\sum_1^n (X_i - \bar{X})^2 + n(\bar{X} - m_0)^2}{\sum_1^n (X_i - \bar{X})^2} > c' \quad \left[\because \sum_1^n (X_i - \bar{X}) = 0\right]
\end{aligned}
$$

এখানে আমরা $(a+b)^2$-এর ফর্মুলা লাগিয়েছি।

$$
\begin{aligned}
\iff \quad & 1 + \frac{n(\bar{X} - m_0)^2}{\sum_1^n (X_i - \bar{X})^2} > c' \\
\iff \quad & \frac{n(\bar{X} - m_0)^2}{\sum_1^n (X_i - \bar{X})^2} > \underbrace{c' - 1}_{c'', \text{ say}} \\
\iff \quad & \frac{\sqrt{n}|\bar{X} - m_0|}{\sqrt{\frac{1}{n-1}\sum_1^n (X_i - \bar{X})^2}} > \underbrace{\sqrt{(n-1)c''}}_{c''', \text{ say.}}
\end{aligned}
$$

এই শেষের ধাপটা করেছি t-distribution-এ পৌঁছনোর জন্য।

> We need to choose c''' such that
>
> $$\alpha,$$
>
> the level of significance.
>
> Step 4: Now, under H_0, we know that
>
> $$\frac{\sqrt{n}(\bar{X}-m_0)}{\frac{1}{n-1}\sum_1^n(X_i-\bar{X})^2} \sim t_{(n-1)}.$$
>
> Let $F(\cdot)$ be the CDF of this distribution.
> Then
>
> $$P\left(\frac{\sqrt{n}|\bar{X}-m_0|}{\sqrt{\frac{1}{n-1}\sum_1^n(X_i-\bar{X})^2}} > c'''\right) = 2(1-F(c''')),$$
>
> since the distribution is symmetric around 0.

এই জিনিসটা কেন হল সেটা ভেবে দ্যাখো। ছবি আঁকলে সুবিধা হবে।

> Thus we need $2(1-F(c''')) = \alpha$,
> or, $c''' = F^{-1}\left(1-\frac{\alpha}{2}\right)$.

■

DAY 39
Likelihood Ratio Tests (approximate)

হিজিবিজ্‌বিজ্‌ বলল, "হ্যাঁ, গানটা ভারি শক্ত।"
ছাগল বলল, "শক্ত আবার কোথায়? ঐ শিশিবোতলের জায়গাটা একটু শক্ত ঠেকল, তাছাড়া তো শক্ত কিচ্ছু পেলাম না।"

--হ য ব র ল (সুকুমার রায়)

গতকাল যে likelihood ratio test-গুলো বার করা শিখেছিলাম, সেগুলো তেমন শক্ত কিছু ছিল না, খালি একটা জিনিস ছাড়া। আমরা H_0-কে reject করছিলাম "$LR < c$" হলে। এই LR-এর চেহারাটা সাধারণতঃ অতিশয় বদখত আসছিল। যেমন normal mean-এর অংকে $LR < c$ মানে ছিল

$$\exp\left(-\frac{\sum_1^n(x_i-m_0)^2}{2\sigma^2}\right)\Big/\exp\left(-\frac{\sum_1^n(x_i-m_1)^2}{2\sigma^2}\right) < c.$$

সেইটাকে মালিশ করে করে আমরা পরিণত করেছিলাম $\sum x_i$-ওয়ালা একটা inequality-তে। এই মালিশটা করার দরকার পড়েছিল কারণ $\sum x_i$-এর sampling distribution জানা আছে, কিন্তু ওই বদখত LR-টার sampling distribution বার করতে গেলে আর রক্ষা ছিল না!

দুঃখের কথা হল, এইভাবে মালিশ করে সবসময়ে LR-কে একটা সহজ কিছু জিনিসে এনে ফেলা যায় না। তখন অগত্যা কিছু approximation-এর শরণাপন্ন হতে হয়। এরকম কিছু উদাহরণ শিখব আজকে।

39.1 Normal variance

Example 31: Using the method of likelihood ratio testing describe a method of testing hypothesis $H_0 : \sigma = \sigma_0$ for a normal(m, σ) population.[6] **(2007.5c)**

SOLUTION: শুরুটা যথারীতি--

Step 0: Let $X_1, ..., X_n$ be IID $N(m, \sigma)$ where $m \in \mathbb{R}$ and $\sigma > 0$ are both unknown. The parameter space is $\Theta = \{(m, \sigma) \ : \ m \in \mathbb{R}, \sigma > 0\}$.

To test

$$H_0 : \sigma = \sigma_0 \text{ against } H_1 : \sigma \neq \sigma_0.$$

Here $\sigma_0 > 0$ is some given number. So $\Theta_0 = \{(m, \sigma_0) \ : \ m \in \mathbb{R}\}$. The likelihood function is

$$\begin{aligned} L(m, \sigma) &= \prod_1^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(X_i - m)^2\right) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_1^n (X_i - m)^2\right) \end{aligned}$$

Step 1: $\sup\{L(m, \sigma) \ : \ (m, \sigma) \in \Theta_0\} = \sup\{L(m, \sigma_0) \ : \ m \in \mathbb{R}\} = L(\widehat{m}_0, \sigma_0)$, where $\widehat{m}_0$ is the MLE for m when $\sigma = \sigma_0$ is known. From standard result we know that $\widehat{m}_0 = \bar{X}$. So

$$\sup\{L(m, \sigma) \ : \ (m, \sigma) \in \Theta_0\} = L(\bar{X}, \sigma_0).$$

Step 2: $\sup\{L(m, \sigma) \ : \ (m, \sigma) \in \Theta\} = \sup\{L(m, \sigma) \ : \ m \in \mathbb{R}, \sigma > 0\} = L(\widehat{m}, \widehat{\sigma})$, where $(\widehat{m}, \widehat{\sigma})$ is the MLE for (m, σ). From standard result we know that $\widehat{m} = \bar{X}$ and $(\widehat{\sigma})^2 = \frac{1}{n}\sum (X_i - \bar{X})^2$. So

$$\sup\{L(m, \sigma) \ : \ (m, \sigma) \in \Theta\} = L\left(\bar{X}, \sqrt{\tfrac{1}{n}\sum (X_i - \bar{X})^2}\right).$$

Step 3: We have

$$\begin{aligned} L(\widehat{m}_0, \sigma_0) &= \prod_1^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(X_i - m)^2\right) \\ &= (2\pi\sigma_0^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_0^2}\sum_1^n (X_i - \bar{X})^2\right) \end{aligned}$$

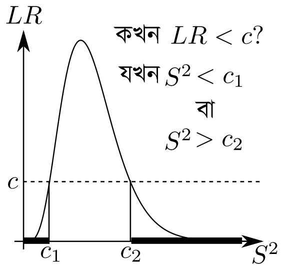


Fig 12

Also

$$
\begin{aligned}
L(\widehat{m}, \widehat{\sigma}) &= \cdots \\
&= (2\pi\widehat{\sigma}^2)^{-n/2} e^{-n/2} \\
&= \left(2\pi \tfrac{1}{n} \sum (X_i - \bar{X})^2\right)^{-n/2} e^{-n/2}.
\end{aligned}
$$

Thus the likelihood ratio is

$$LR = \frac{L(\widehat{m}_0, \sigma_0)}{L(\widehat{m}, \widehat{\sigma})} = e^{n/2}(S^2)^{n/2} e^{-nS^2/(2\sigma_0^2)},$$

where $S^2 = \frac{1}{n}\sum(X_i - \bar{X})^2$ is the sample variance.

এবার দেখতে হবে $LR < c$ মানে কী। এখানে LR-টা এতই বিচ্ছিরি দেখতে যে একটু গ্রাফ এঁকে নিলে সুবিধা হবে। গ্রাফ আঁকাটা খুব কঠিন নয়, কারণ জিনিসটা কতকটা gamma function-এ যেটাকে integrate করা হয় সেটার মত। তার চেহারাটা চেনা থাকা উচিত। নেহাত যদি না থাকে Fig 12 থেকে দেখে নাও।

The graph of this ratio as a function of S^2 looks like Fig 12. From this we see that H_0 is to be rejected both when S^2 is very small or very large.

Step 4: The exact cut off values are difficult to be computed. So approximately, we reject H_0 if and only if $S^2 < c_1$ or $S^2 > c_2$, where $P(S^2 < c_1) = P(S^2 > c_2) = \frac{\alpha}{2}$ under H_0.

By standard result, $\frac{n}{\sigma_0^2} S^2 \sim \chi^2_{(n-1)}$ under H_0.

Let $F(t)$ be the cdf of this distribution.

Then $P\left(\frac{n}{\sigma_0^2} S^2 < F^{-1}(\alpha/2)\right) = \alpha/2$ and $P(\frac{n}{\sigma_0^2} S^2 > F^{-1}(1-\alpha/2)) = \alpha/2$.

Hence we take $c_1 = \sigma_0^2 F^{-1}(\alpha/2))/n$ and $c_2 = \sigma_0^2 F^{-1}(1-\alpha/2))/n$.

■

এবার একটা প্রয়োগ দেখি।

Example 32: The variance of scores on a standard mathematics test for all high school students

of Calcutta was 50 in 2000. A random sample of scores of 20 high school students who took the test this year gave a variance of 70. Test at 5% level of significance if the variance of the current scores on this test differ from 50. [Given $\chi^2_{0.975} = 8.97$ and $\chi^2_{0.025} = 32.8$].[6] **(2014.6a)**

SOLUTION: এই অংকটা দেখে তোমাকে চিনতে হবে যে, এটা আসলে আগের অংকটারই প্রয়োগ। আগের অংকে নানারকম assumption দেওয়া ছিল, যেমন normal distribution ইত্যাদি। এখানে সে সব কিছুই দেওয়া নেই। সেগুলো তোমাকে লিখে নিতে হবে। তারপর আগের অংকে যে test-টা পেয়েছিলে সেটাকে সরাসরি ব্যবহার করতে হবে (নতুন করে আর বার করতে হবে না)।

সুতরাং শুরু করব assumption-গুলো বলে--

> Let the scores of the $n = 20$ sample students be $X_1, ..., X_n$.
> We assume these to be iid $N(\mu, \sigma)$, where $\mu \in \mathbb{R}$ and $\sigma > 0$ are both unknown.

এইখানে তোমার মনে খুবই সঙ্গত একটা প্রশ্ন উঠতে পারে-- কী করে জানলাম যে, ছাত্রদের পরীক্ষার নম্বরগুলোর এই distribution হবে? উত্তর হল--"কারণ ওই একটা distribution-এর জন্যই আমরা variance-এর test করতে শিখেছি!" এটা অবশ্য নিতান্তই পরীক্ষা-মার্কা হাস্যকর উত্তর হল। বাস্তবে এরকম অন্ধের মত distribution কল্পনা করে নিলে ভুলভাল জিনিস বেরোবে। আসল কায়দাটা হওয়া উচিত এইরকম--প্রথমে আমাদেরকে sample-এর histogram দেখে আন্দাজ করতে হবে distribution-টা কী, তারপর সেই distribution-এর ভিত্তিতে test-টা তৈরী করতে হবে।

তারপর কোথ্‌থেকে একটা পুরোনো দরজির ফিতে এনে সে আমার মাপ নিতে শুরু করল, আর হাঁকতে লাগল "খাড়াই ছাব্বিশ ইঞ্চি, হাতা ছাব্বিশ ইঞ্চি, আস্তিন ছাব্বিশ ইঞ্চি, ছাতি ছাব্বিশ ইঞ্চি, গলা ছাব্বিশ ইঞ্চি।"
আমি ভয়ানক আপত্তি করে বললাম, এ হতেই পারে না, বুকের মাপও ছাব্বিশ ইঞ্চি, গলাও ছাব্বিশ ইঞ্চি? আমি কি শুয়োর?
বুড়ো বলল, বিশ্বাস না হয়, দেখ।
দেখলাম ফিতের লেখা-টেখা সব উঠে গিয়েছে, খালি ছাব্বিশ লেখাটা একটু পড়া যাচ্ছে, তাই বুড়ো যাই মাপে সবই ছাব্বিশ ইঞ্চি হয়ে যায়।

--হ য ব র ল (সুকুমার রায়)

যাই হোক, এখানে আমরা সেই পরীক্ষা-মার্কা পথেই এগোব--

> Our approximate LRT rejects H_0 at 5% level of significance if the sample variance $S^2 = \frac{1}{n}\sum_1^n (X_i - \bar{X})^2$ lies outside $[c_1, c_2]$, where $P(S^2 < c_1) = P(S^2 > c_2) = \frac{\alpha}{2}$ under H_0.
>
> By standard result,
>
> $\frac{nS^2}{\sigma_0^2} \equiv \frac{1}{\sigma_0^2}\sum_1^n (X_i - \bar{X})^2 \sim \chi^2_{(n-1)}$ under H_0.
>
> Let $F(t)$ be the cdf of this distribution.
>
> Then under H_0, we have
>
> $P(S^2 < \frac{\sigma_0^2}{n} F^{-1}(\alpha/2)) = \alpha/2$ and $P(S^2 > \frac{\sigma_0^2}{n} F^{-1}(1-\alpha/2)) = \alpha/2$
>
> Hence we take $c_1 = \frac{\sigma_0^2}{n} F^{-1}(\alpha/2))$ and $c_2 = \frac{\sigma_0^2}{n} F^{-1}(1-\alpha/2))$.
>
> For the given values we get $c_1 = 22.425$ and $c_2 = 82.000$.
>
> Thus we reject H_0 if S^2 lies outside $[22.425, 82.000]$.
>
> Since for the given data $S^2 = 70 \in [c_1, c_2]$, we accept H_0 at 5% level of significance to conclude that the given data do not provide significant

evidence against believing that the variance of the current scores is 50.

■

LRT-র কায়দা ব্যবহার করে যে সব test পাওয়া যায়, সেগুলো অনেক সময়েই সহজ বুদ্ধিতেই আন্দাজ করা যায়। যেমন এই যে test-টা করলাম এক্ষুণি, normal variance-এর ব্যাপারে, তার চেহারাটা কী সহজ বুদ্ধিতেই আন্দাজ করা যেত না? আমাদের বলেছিল $\sigma^2 = 50$ কিনা পরীক্ষা করতে। এখানে σ^2 আমরা জানি না, তাই তার বদলে বার করলাম sample variance, মানে $\frac{1}{n}\sum(X_i - \bar{X})^2$. এবং দেখলাম এই জিনিসটাকে 50-এর সঙ্গে তুলনা করেছিলাম, বা ঘুরিয়ে বললে $\frac{1}{n}\sum(X_i - \bar{X})^2/50$-কে 1-এর সঙ্গে তুলনা করেছিলাম। যদি 1-এর থেকে যথেষ্ট দূরে থাকে, তবে H_0-কে reject করছিলাম। Critical value-গুলো বার করতে খানিকটা অংক করতে হয়েছিল, ঠিকই, কিন্তু test-টার মূল চেহারাটা সহজ বুদ্ধিতেই আন্দাজ করা যাচ্ছিল, অত likelihood-এর sup-টুপ বার না করেই।
এই ব্যাপারটা অনেক সময়েই দেখা যায়--LRT যদি একেবারে সংজ্ঞা মেনে likelihood-এর sup বার করে ধুন্ধুমার করে বানাও, তবে অনেক খাটনি যায়। অথচ সহজ বুদ্ধিতে চিন্তা করলে অনেক সময়ে এমনিতেই একই test-এ পৌঁছনো যায়। তাই সহজ বুদ্ধির কায়দাটাও জেনে রাখা ভালো। নীচের অংকদুটো তোমাকে সেই তালিমটাই দেবে।

39.2 Testing proportion

Example 33: Design a decision rule to test the hypothesis that a coin is fair, if a sample of 64 tosses of the coin is taken and if a level of significance of 0.05 is used. Given that

$$\frac{1}{\sqrt{2\pi}}\int_0^{1.96} e^{-x^2/2}dx = 0.4770.$$

[5] **(2009.5c)**
SOLUTION: প্রথমেই লক্ষ কর যে, এখানে কিন্তু LRT বার করতে বলে নি মোটেই। কিন্তু আসলে যেটা বার করব সেটা LRT-ই হবে (যদিও approximate)। তবে সরাসরি LRT-র সংজ্ঞা লাগিয়ে এগোতে গেলে কেমন ঝকমারি হয় দ্যাখো--এখানে 64 বার টস করা হয়েছে, সুতরাং data হল $X_1, ..., X_{64}$. একটু চিন্তা করলেই বুঝবে যে likelihood-টা হবে

$$L(\theta) = \theta^X(1-\theta)^{64-X},$$

যেখানে X হল মোট head-এর সংখ্যা।
আমাদের hypothesis-রা হল $H_0 : \theta = \frac{1}{2}$ বনাম $H_1 : \theta \neq \frac{1}{2}$. যেহেতু H_0-টা এখানে simple hypothesis, তাই LR-এর উপরতলার sup-টা সহজ, স্রেফ $L\left(\frac{1}{2}\right)$. নীচের তলার sup-টা হল $L(\widehat{p})$, যেখানে $\widehat{p} = \frac{X}{64}$ হল p-এর MLE. সুতরাং LR-টা দাঁড়াচ্ছে--

$$LR = \frac{\sup_{\theta\in\Theta_0} L(\theta)}{\sup_{\theta\in\Theta} L(\theta)} = \frac{32^{64}}{X^X(64-X)^{64-X}}.$$

এই অতিশয় বিচ্ছিরি দেখতে পদার্থটাকে মালিশ করে একটা ভদ্র চেহারা দেওয়া সম্ভব বটে, কিন্তু সেটা বেশ মেহনতের ব্যাপার। তার চেয়ে আমরা একই test-এ পৌঁছতে পারি খালি সহজ বুদ্ধি দিয়েই। কী করে দেখা যাক।

Step 0: Let $X_1, ..., X_{64}$ be the outcomes of the 64 tosses, where a head is considered as 1, and a tail as 0.

Then X_i's are IID Bernoulli(θ), where $p = P(\text{head})$ is the unknown parameter and $\Theta = (0, 1)$ is the parameter space.

> To test
> $$H_0 : p = \frac{1}{2} \text{ against } H_1 : p \neq \frac{1}{2}.$$
> Thus $\Theta_0 = \left\{\frac{1}{2}\right\}$.
> We know that $\widehat{p} = \frac{X}{n}$ lies close to p, where X is the total number of heads.

সুতরাং $\widehat{p}$-কে p-এর প্রক্সি হিসেবে ব্যবহার করব। কিন্তু $\widehat{p}$ তো একেবারে p-এর সমান নাই হতে পারে, তাই $\widehat{p} \neq \frac{1}{2}$ দেখলেই অমনি জোর দিয়ে $p \neq \frac{1}{2}$ বলা যাবে না। কিন্তু যেহেতু p এবং $\widehat{p}$ কাছাকাছি থাকাটা প্রত্যাশিত, তাই যদি দেখি $\widehat{p}$-টা $\frac{1}{2}$-এর চেয়ে বেশ অনেকটাই দূরে রয়েছে, তবে আশা করা অসঙ্গত হবে না যে, $p \neq \frac{1}{2}$ হবে।

> So it is reasonable to reject H_0 if and only if
> $$|[|\widehat{p} - \tfrac{1}{2}|]| > c$$
> for some c such that the level of the test is 5%, ie, under H_0
> $$P\left(|[|\widehat{p} - \tfrac{1}{2}|]| > c\right) = 0.05.$$

মজার কথা হল ঠিক এই একই test-এ আমরা পৌঁছতাম LRT-র সংজ্ঞা ব্যবহার করেও। খালি সে পথে গেলে ধাপগুলো অনেক বেশী লম্বা আর জটিল হত। এবার আমাদের কাজ হল c বার করা, এবং সেই কাজে approximation লাগাতে হবে।

> Step 2: By CLT that the distribution of
> $$\frac{\widehat{p} - p}{\sqrt{p(1-p)/n}}$$
> is asymptotically $N(0,1)$.

যদি H_0 ঠিক হয়, তবে $p = \frac{1}{2}$. সুতরাং--

> So, under H_0, the distribution of
> $$2\sqrt{n}\left(\widehat{p} - \tfrac{1}{2}\right)$$
> is approximately $N(0,1)$. So
> $$P\left(|[|\widehat{p} - \tfrac{1}{2}|]| > c\right) = P\left(2\sqrt{n}|[|\widehat{p} - \tfrac{1}{2}|]| > 2\sqrt{n}c\right) = 2(1 - \Phi(2\sqrt{n}c)),$$
> where $\Phi(\cdot)$ is the $N(0,1)$ CDF.
> So we need $2(1 - \Phi(2\sqrt{n}c)) = 0.05$,
> or $c = \frac{\Phi^{-1}(0.975)}{2\sqrt{n}} = 0.123$.
> Thus our test is to reject H_0 if and only if $|[|\widehat{p} - \frac{1}{2}|]| > 0.123$.

■

এবারের একইরকম আরেকটা অংক দেখব। এখানেও LRT-র সংজ্ঞা সরাসরি লাগালে ব্যাপারটা বেজায় জটিল হয়ে উঠবে, কিন্তু সহজ বুদ্ধিতে একই test-এ পৌঁছোনো সম্ভব। আমরা সহজ বুদ্ধির পথেই এগোব।

Example 34: In random samples of 374 and 210 from the adult populations of two large cities

72.4% and 88.1% were respectively found to be literate. Do the two populations really differ in their percentage of literacy? [Given $\frac{1}{\sqrt{2\pi}}\int_{4.36}^{\infty} e^{-x^2/2}dx < 0.0001$.][5] **(2010.5d)**

SOLUTION: অংকটা কী চাইছে বুঝে নিই আগে। এখানে দুটো শহর আছে, যাদের জনসংখ্যা অনেক। প্রত্যেকটা মানুষই হয় সাক্ষর নয়তো নিরক্ষর। বোঝার সুবিধার জন্য ধরো প্রথম শহরে সাক্ষরতার হার হল p_1, আর দ্বিতীয় শহরে p_2. বাস্তবে এই হারদুটো জানা খুবই কঠিন, কারণ তার জন্য প্রতিটা লোককে ধরে ধরে পরীক্ষা করে দেখতে হবে সাক্ষর নাকি নিরক্ষর। যেহেতু জনসংখ্যা বলা আছে অনেক, তাই সে কাজটা একরকম অসম্ভব বললেই হয়। অথচ সাক্ষরতার হার জানাটাও অনেক সময়েই দরকার পড়ে (যেমন ধরো কেন্দ্র থেকে শিক্ষাখাতে বরাদ্দ টাকা দুই শহরের মধ্যে কীভাবে ভাগ করা উচিত তা স্থির করার জন্য)। সেই জন্য দুই শহর থেকেই কিছু লোকের একটা করে random sample নেওয়া হয়। এখানে যেমন প্রথম শহর থেকে 374 জন আর দ্বিতীয় শহর থেকে 210 জনকে নেওয়া হয়েছে। এদের প্রত্যেককে ধরে ধরে পরীক্ষা করে দেখা হয় কতজন সাক্ষর। তাতে দেখা গেল যে প্রথম শহরের sample-এ সাক্ষরতার হার হল $\widehat{p}_1 = 72.4\%$, আর দ্বিতীয় শহরের sample-এ হারটা হল $\widehat{p}_2 = 88.1\%$. মনে রেখো এইগুলো হল sample-এর ভিত্তিতে পাওয়া সাক্ষরতার হার। যেহেতু আস্ত শহরটার সবগুলো লোককে ধরে ধরে আমরা পরীক্ষা করিনি, তাই এই হারগুলো পুরো শহরের সাক্ষরতার হারের একেবারে সমান হবে এমনটা জোর দিয়ে বলা যায় না। কিন্তু কাছাকাছি কিছু একটা হবে এমনটা আশা করা অসঙ্গত নয়। মানে $p_1 \approx \widehat{p}_1$ এবং $p_2 \approx \widehat{p}_2$. আমাদের testকরতে বলেছে $H_0 : p_1 = p_2$ বনাম $H_1 : p_1 \neq p_2$.
এখানে সহজ বুদ্ধিতে এগোনোর কায়দাটা হল $\widehat{p}_1$ আর $\widehat{p}_2$-কে যথাক্রমে p_1 আর p_2-র প্রক্সি হিসেবে ব্যবহার করা। কিন্তু যেহেতু p_1 আর $\widehat{p}_1$ (এবং p_2 আর $\widehat{p}_2$) একেবারে সমান নাও হতে পারে, তাই $\widehat{p}_1 \neq \widehat{p}_2$ হলেই অমনি জোর দিয়ে বলা যাবে না যে, $p_1 \neq p_2$ হবেই। কিন্তু যদি $\widehat{p}_1$আর $\widehat{p}_2$ পরস্পরের থেকে অনেকটাই দূরে থাকে, তবে সিদ্ধান্ত করা যেতেই পারে যে p_1 নিশ্চয়ই p_2-র থেকে আলাদা। এই যুক্তিটাই এবার গুছিয়ে লিখব।

> Step 0: Let the population proportions be p_1 and p_2.
> To test
> $$H_0 : p_1 = p_2 \text{ against} H_1 : p_1 \neq p_2.$$
> Here the sample sizes are $n_1 = 374$ and $n_2 = 210$. The sample proportions are $\widehat{p}_1 = 0.724$ and $\widehat{p}_2 = 0.881$, respectively.
> Step 1: Since we can expect that $\widehat{p}_1 \approx p_1$and $\widehat{p}_2 \approx p_2$ so we shall reject H_0 if $|\widehat{p}_1 - \widehat{p}_2| > c$, for some c such that the level of the test is 5%, ie, $P(|\widehat{p}_1 - \widehat{p}_2| > c) = 0.05$ under H_0.

এবার c বার করার পালা। তার জন্য দেখতে হবে H_0 ঠিক হলে $\widehat{p}_1 - \widehat{p}_2$-এর distribution কী হয়। একাজে আমরা approximation লাগাব।

> Step 2: For $i = 1, 2$, we know that $\frac{\widehat{p}_i - p_i}{\sqrt{p_i(1-p_i)/n_i}}$ is asymptotically $N(0, 1)$ by CLT.

So the distribution of $\widehat{p}_i$ is approximately

$$N\left(p_i, \sqrt{\frac{p_i(1-p_i)}{n_i}}\right).$$

Assuming the two samples to be independent, the distribution of $\widehat{p}_1 - \widehat{p}_2$ is approximately

$$N\left(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right).$$

আমাদের দরকার খালি null distribution-টা, তাই--

Under $H_0 : p_1 = p_2 (= p,$ say) this is

$$N\left(0, \underbrace{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}_{\sigma}\right).$$

So the distribution of

$$\frac{\widehat{p}_1 - \widehat{p}_2}{\sigma}$$

is approximately $N(0, 1)$.

দুঃখের ব্যাপার হল p-টাও আমরা জানি না। তাই সেটাকে estimate করতে হবে, ফলে আরেক প্রস্থ approximation ঢুকবে--

If the two populations have the same proportion (p, say), then it can be estimated by $\widehat{p} = \frac{n_1\widehat{p}_1 + n_2\widehat{p}_2}{n_1+n_2} = 0.780$.

Now $\widehat{p} \approx p$, and so

$$\widehat{\sigma} = \sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \approx \sigma.$$

So the distribution of

$$\frac{\widehat{p}_1 - \widehat{p}_2}{\widehat{\sigma}}$$

is also approximately $N(0, 1)$. Hence we can reject H_0 if

$$||[|\frac{\widehat{p}_1 - \widehat{p}_2}{\widehat{\sigma}}|]|| > \Phi^{-1}(0.975) = 1.96.$$

> In our case the LHS is 0.036. So we accept H_0 to conclude that the given data do not provide significant evidence at 5% level against believing that the two population proportions are the same.

■

39.3 Statistical regularity (আবার!)

Likelihood ratio test খুবই জনপ্রিয় একটা কায়দা। কিন্তু খুব কম ক্ষেত্রেই LR-টাকে মালিশ করে এমন কোনো কিছুতে পরিণত করা যায় যার sampling distribution জানা আছে। Approximation-এর যে কয়টা উদাহরণ আমরা দেখলাম, তারাও মোটেই general নয়। এই উদাহরণগুলোতে সব ক্ষেত্রেই একটা করে প্রক্সি পাওয়া যাচ্ছিল, এবং চট্ করে CLT-ও লাগানো যাচ্ছিল। সবসময়ে সে সুবিধা থাকে না। General কেসের জন্য একটা approximation আছে যেটার সম্পূর্ণ আলোচনা এই বইয়ের স্বল্পপরিসরে সম্ভব নয়, কিন্তু একটু উল্লেখ করে রাখি। কায়দাটা হল এই--যদি sample size যথেষ্ট বড় হয় তবে H_0 ঠিক হলে $-2\log(LR)$-র distribution-টা একটা χ^2-distribution-এর দিকে এগিয়ে যায়। এর জন্য অবশ্য কিছু শর্ত লাগে, কিন্তু সে অতি সামান্যই। সেই সব শর্ত কী, বা এই χ^2-এর degrees of freedom-ই বা কত, সেইসব প্রশ্নের মধ্যে যাব না। খালি বলে রাখি আমরা কিছুদিন আগে যে χ^2 goodness-of-fit test শিখেছিলাম, সেটাও আসলে এরই প্রয়োগ। এবং এই জিনিসটার মূলে আছে সেই statistical regularity-র ধারণাটাই।

তাহলেই বোঝো statistical regularity আমাদের কীরকম নিত্যসঙ্গী। সেই প্রথম অধ্যায়ের প্রথম দিনেই এর দেখা মিলেছিল। আজ আবার তার হাতের উপহার নিয়েই এ বইয়ের ইতি টানলাম।

Answers

1. Data হল টস দুটো outcome যাদের নাম দিতে পারি X_1, X_2. মোট head-এর সংখ্যা হল test statistic. আর critical region হচ্ছে $\{HH\}$, মানে যখন X_1, X_2 দুজনেই head. **3.** $P(\text{type I error}) = 0.1875$ আর $P(\text{type II error}) = 0.3672$. **5.** $P(\text{type I error}) = 0.5$ $P(\text{type II error}) = 0.25$.

Index